THE SEARCH FOR THE PFHBI GENE: REFINING THE TARGET AREA AND IDENTIFICATION AND ANALYSIS OF CANDIDATE GENE TRANSCRIPTS

Zainunisha Arieff

Dissertation presented for the degree of Doctor of Philosophy at the Faculty of Health Sciences, University of Stellenbosch.

Promoter: Professor Valerie Corfield

December 2004

DECLARATION

I, the undersigned hereby declare that the work contained in this dissertation is my own original work and that I have not previously in its entirety or in part submitted it at any university for a degree.

Signature: **Date:**

SUMMARY

Progressive familial heart block I (PFHBI) is an inherited autosomal dominant cardiac conduction disorder which segregates in a large South African (SA) pedigree, two smaller SA families and a Lebanese family. It specifically affects conduction in the ventricles and is of unknown cause. Clinically, PFHBI is detected on electrocardiogram (ECG) by evidence of bundle-branch disease, i.e., as right bundle branch block, left anterior or posterior hemiblock, or complete heart block with broad QRS complexes. The PFHBI-causative gene was mapped to a 10cM region on chromosome 19q13.3 using linkage analysis, and the locus was subsequently reduced to 7cM by genetic fine mapping.

The present study involved a multi-strategy approach to search for the PFHBI gene. The objectives were the further reduction of the PFHBI locus by genetic fine mapping using published and novel markers, searching for short gene transcripts from publicly available databases and the generation of an integrated map of the locus to which genes were mapped. Prioritised genes were screened for PFHBI-causing mutations and, in addition, the PFHBI locus was searched for the presence of a G protein-encoding gene (*P115-RhoGEF*), a connexin (*Cx*) gene and any genes containing a CTG repeat expansion motif, since these genes are plausible PFHBI candidate genes.

Genotyping and fine genetic mapping using known and novel polymorphic dinucleotide $(CA)_n$ and novel tetranucleotide $(A_3G)_n$ repeat markers across the PFHBI locus were performed. Publicly available databases, such as LLNL (Livermore, USA), and GENEMAP (NCBI) were searched for ESTs which, in turn, were extended using clustering programmes, such as UNIGENE (NCBI) and STACK (SANBI), and the resulting consensus sequences were subsequently BLAST-searched against the protein databases. Using the available data, an integrated physical and genetic map of the PFHBI locus was generated and, as the HGP progressed, a number of novel genes were placed thereon. Subsequently, genes were prioritised on the basis of position, function and

expression profile.

Genetic fine mapping reduced the PFHBI locus from 7cM to 4cM. The EST approach yielded 38 ESTs, of which 24 ESTs matched proteins, such as activating transcription factor 5 (*ATF5)*, actin-binding protein *(KPTN)* and zinc finger protein 473 (*ZFP473*) (May 2003). All the map data generated experimentally and computationally were placed on the PFHBI map. The *P115-RhoGEF* was excluded as a PFHBI candidate gene and although homologous sequences to connexin 37 (*Cx37)* was located on both chromosome 19 radiation hybrid clones (RHG12 and ORIM-7), it was not identified on the DNA clones spanning the PFHBI locus. No evidence of an expansion of a CTG repeat motif sequence in PFHBI-affected individuals was found. Five highly prioritised candidate genes, namely, *BCL2*-associated X protein (*BAX)*, potassium voltage-gated channel Shaker-related subfamily member 7 (*KCNA7)*, potassium inwardly-rectifying channel, subfamily J, member 14 (*KIR2.4*), lin-7 homolog B (*LIN-7B)* and glycogen synthase 1 (*GSYI*) were selected for mutation screening. No disease associated mutations were identified in the exonic and flanking intronic regions of these genes.

In summary, this study reduced the PFHBI locus substantially and generated a detailed map of the region. A number of attractive candidate genes were excluded from causing PFHBI; however, several plausible candidate genes are still present at this gene-rich locus and remain to be screened. Identifying the PFHBI-causative gene and associated mutation will provide a platform for further studies to understand the pathophysiology, not only of PFHBI, but also of other more commonly occurring conduction disturbances.

OPSOMMING

Progressiewe familiële hartblok I (PFHBI) is 'n autosomaal dominant oorerflike kardiale geleidingstoornis wat in 'n groot Suid-Afrikaanse (SA) familie, twee kleiner SA families en 'n Lebanese familie segregeer. Dit affekteer hoofsaaklik die geleiding in die ventrikels en die oorsaak daarvan is onbekend. Klinies word PFHBI op elektrokardiogram (EKG) geïdentifiseer as a bondeltak-siekte, naamlik, as regter bondeltakblok, linker anterior of posterior hemiblok, of volledige hartblok met wye QRS komplekse. Die PFHBI-veroorsakende geen is voorheen deur koppelingsanalise tot 'n 10cM gebied op chromosoom 19q13.3 gekarteer, en daaropvolgens is die lokus verklein tot 7cM deur genetiese fyn kartering.

Die huidige studie behels 'n veelvuldige-strategie benadering in die soektog na die PFHBI geen. Die doel van die studie was die verdere verkleining van die PFHBI lokus deur gebruik te maak van beide gepubliseerde en nuwe genetiese merkers, die identifisering van kort geentranskripte (ESTs) uit publieke databanke en die generasie van 'n geïntegreerde kaart van die lokus. Geprioritiseerde gene is geanaliseer vir die PFHBI-veroorsakende mutasie en, daarby, is die PFHBI lokus deursoek vir die teenwoordigheid van 'n G proteïen-enkodeeringsgeen (*P115-RhoGEF*), 'n konneksien (Kx) geen en enige gene wat 'n uitgebreide CTG-herhalingsmotief bevat, aangesien hierdie gene as sterk PFHBI kandidaatgene geag is.

Genotipering en fynkartering deur die gebruik van bekende asook nuwe polimorfiese dinukleotied- [(CA)n] en nuwe tertranukleotied- [(A_3G)n] herhalingsmerkers wat die PFHBI lokus oorbrug, is uitgevoer. Publieke databanke, soos LLNL (Livermore, USA), en GENEMAP (NCBI) is ondersoek vir ESTs wat vervolgens verleng is deur gebruik te maak van groeperende programme soos UNIGENE (NCBI) en STACK (SANBI) en die gevolglike konsensus volgordes is daarna met behulp van BLAST geanaliseer teen die proteïendatabanke. Die bekomde data is vervolgens

gebruik om 'n geïntegreerde fisiese en genetiese kaart van die PFHBI lokus te produseer en, soos die mens genoomprojek gevorder het, is nuwe gene daarop geplaas. Daarna is gene geprioritiseer vir mutasie analise gebaseer op posisie, funksie en uitdrukkingsprofiele.

Genetiese fynkartering het die PFHBI lokus van 7cM tot 4cM verklein. Die EST benadering het 38 ESTs geïdentifiseer, waarvan 24 ESTs proteïen gelyke gehad het, bv aktiverende transkripsie faktor 5 (*ATF5*), aktien-verbindingsproteïen (*KPTN*) en sink-vingerproteïen 473 (*ZFP473*) (Mei 2003). Al die karterings data wat eksperimenteel en rekenaar-gewys gegenereer is, is op die PFHBI kaart geposisioneer. Die *P115-RhoGEF* is uitgeskakel as 'n PFHBI kandidaatgeen en alhoewel 'n volgorde met homologie aan konneksien37 (Kx37) gevind is op albei chromosoom 19 radiasie-hibried klone (RGH12 and ORIM-7), is dit nie geïdentifiseer in die DNS klone wat die PFHBI lokus oorbrug nie. Geen bewyse van uitbreiding van CTG herhalingsmotiewe is gevind in PFHBI-aangetasde persone nie. Vyf hoogs-geprioritiseerde kandidaat gene, naamlik, BCL2-geassosieerde X proteïen (*BAX*), kalium spanningsbeheerde kanaal, subfamilie J, lid 14 (*KIR2.4*), lin-7 homoloog B (*LIN-7b*) en glikogeen sintase 1 (*GYS1*), is geselekteer vir mutasie-analise. Geen siekte-veroorsakende mutasie is egter geïdentifiseer in die eksoniese of die naasliggende introniese gebiede van hierdie gene nie.

Ter opsomming, hierdie studie het die PFHBI lokus verklein en het 'n omvattende kaart van die gebied gegenereer. Verskillende kandidaat gene is uitgesluit as die oorsaak van PFHBI, alhoewel daar nog heelwat goeie kandidaat gene in hierdie geen-ryke lokus is wat geanaliseer behoort te word. Die identifiseering van die PFHBI-veroorsakende mutasie sal 'n platform bied vir verdere studies om die patofisiologie van nie alleen PFHBI nie, maar ook meer algemene geleidingstoornisse, te verstaan.

TEACHING

Then said a teacher, Speak to us of Teaching.

And he said:

No man can reveal to you aught, but that which already lies half asleep in the dawning of your knowledge.

The teacher who walks in the shadow of the temple, among his followers, gives not of his wisdom but rather of his faith and his lovingness.

If he is indeed wise he does not bid you enter the house of his wisdom, but rather leads you to the threshold of your own mind.

The astronomer may speak to you of his understanding of space, but he cannot give you his understanding of space,

The musician may sing to you of rhythm, which is in all space, but he cannot give you the ear, which arrests the rhythm, nor the voice that echoes it.

And he who is versed in the science of genetics can tell of the regions of the chromosomes, but he cannot cure diseases.

For the vision of one man lends not its wings to another.

And even as each one of you stands alone in God's knowledge, so must each of you be alone in his knowledge of God and in his understanding of the earth.

revised from "The Prophet" by Kahlil Gibran

INDEX

ACKNOWLEDGEMENTS

I would like to express sincere gratitude to the following people and institutions and the roles they have played in this study.

To my promoter Valerie Corfield, special thanks for all her constructive criticism, enthusiasm and support throughout this study.

To Paul Brink and past members of the team for their contribution to the PFHBI project.

To Althea Goosen and Ina le Roux for the locating of family members and for DNA extractions, respectively, and also for their friendly encouragement.

Thank you Lab 445, especially Hanlie, Glenda, Pedro and Toy for their support and advice.

I would also like to thank the PFHBI group based at UWC for their teamwork. During the course of the project, the following postgraduate students, Faghri February, B.Sc Hons and M.Sc; Morne Du Plessis, B.Sc Hons and M.Sc (*cum laude*); Alvina Cloete, B.Sc Hons and Michelle McCabe, B.Sc Hons, completed their degrees under my supervision.

To the Wellcome Trust, the NRF and the UWC Research Committee for their financial support for the duration of this project.

To all those wonderful and encouraging friends that helped with the proof reading and kept me motivated and focused.

Finally, a special thank you to my family for being patient, supportive and giving me the opportunity to grow, grow.....

LIST OF ABBREVIATIONS

α	alpha
adRP	autosomal dominant retinitis pigmentosa
$AgNO_3$	silver nitrate
Amp	ampicillin
APS	ammonium persulphate
ASD	atrial septal defect
ATP	adenosine triphosphate
ATPase	adenosine triphosphatase
A-V	atrio-ventricular
BAC	bacterial artificial chromosome
β	beta
BSA	bovine serum albumen
BB	bundle branch
BBB	bundle branch block
°C	degree celsius
Ca^{2+}	calcium ion
$CaCl_2$	calcium chloride
CCS	cardiac conduction system
CED	Camurati-Engelmann disease
CEPH	*Centre d' Études du Polymorphisme Humaine*
CHB	complete heart block
CHLC	Cooperative Human Linkage Centre
Cl^-	chloride ion
cM	centiMorgan
COL	collagen
Cx	connexin
dATP	deoxyadenosine triphosphate
dCTP	deoxycytosine triphosphate
dTTP	deoxythymidine triphosphate
dGTP	deoxyguanine triphosphate
dNTP	deoxynucleotide triphosphate
DDBJ	DNA database of Japan
DM	myotonic dystrophy
DNA	deoxyribonucleic acid
D	aspartic acid

ECG	electrocardiogram
EDTA	ethylenedinitrotetraacetic acid
EMBL	European Molecular Biology Laboratory
EMBO	European Molecular Biology Organisation
EST	expressed sequence tag
ET	endothelin
EtBr	ethidium bromide
γ	gamma
FISH	fluorescence *in situ* hybridisation
FRDA	Friedreich's ataxia
FTP	file transfer protocol
GDP	guanosine diphosphate
GEF	guanine nucleotide exchange factor
GTP	guanosine triphosphate
HGP	Human Genome Project
HGMP	Human Genome Mapping Project
H_2O	water
HOS	Holt-Oram syndrome
hr	hour
http	hypertext transmission protocol
htg	high throughput
ICCD	isolated cardiac conduction disease
IHGSC	International Human genome sequencing consortium
IMAGE	Integrated Molecular Analysis of Gene Expression
JGI	Joint Genome Institute
kb	kilobases
KWE	keratolytic winter erythema
Kn	kanamycin
K^+	potassium ion
LAHB	left anterior hemiblock
LBA	Luria-Bertani agar
LBB	left bundle branch
LBBB	left bundle branch block
LB	Luria-Bertani
LPHB	left posterior hemiblock
LGMD	limb-girdle muscular dystrophy
LLNL	Lawrence Livermore National Laboratory

LOD	logarithm of the odds ratio
LQT	long QT syndrome
M	molar
Mb	megabase
$MgCl_2$	magnesium chloride
min	minute
MLC	myosin light chain
ml	millilitre
Mm	millimolar
Mv	millivolts
mRNA	messenger RNA
Na^+	sodium
NaOH	sodium hydroxide
NCBI	National Centre for Biotechnology Information
ng	nanogram
NIDDM	noninsulin-dependent diabetes mellitus
nm	nanometer
nr	nonredundant
nt	nucleotide
OD	optical density
OMIM	Online Mendelian Inheritance in Man
PAC	P1 artificial chromosome
PCR	polymerase chain reaction
PCCD	progressive cardiac conduction disease
PDZ	postsynaptic density-95, discs large, zonula occludens
PFHBI	progressive familial heart block I
R	arginine
RBB	right bundle branch
RBBB	right bundle branch block
RED	repeat expansion detection
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
RP	retinitis pigmentosa
rpm	revolutions per minute
S	serine
SA	South African

S-A	sino-atrial
SANBI	South African National Bioinformatics Institute
SDS	sodium dodecyl sulphate
sec	second
SLE	systemic lupus erythematosus
SNP	single nucleotide polmorphism
SR	serine, arginine
STACK	Sequence Tag Alignment Consensus Knowledgebase
STR	short tandem repeat
STS	sequence tagged site
SSC	sodium, saline and citrate solution
SSCP	single-strand conformation polymorphism
TBE	tris, boric acid and EDTA buffer
TE	tris, EDTA buffer
Tet	tetracycline
TIGR	The Institute for Genomic Research
Ta	annealing temperature
Tm	melting temperature
µCi	microCurie
µg	microgram
µl	microlitre
µM	micromolar
UK	United Kingdom
US	University of Stellenbosch
USA	United States of America
UTR	untranslated region
vol	volume
WWW	World Wide Web
YAC	yeast artificial chromosome

LIST OF FIGURES

LIST OF TABLES

INTERNATIONAL CODES FOR GENES

AKT1S1	AKT1 substrate 1 (proline-rich)
ANF	atrial natriuretic factor
AP2A1	adaptor-related protein complex 2, alpha 1 subunit
APOB	apolipoprotein B
ATF5	activating transcription factor 5
BAX	*BCL2*-associated X protein
BCAT2	branched chain aminotransferase 2 mitochondrial
BCL2L12	*BCL2*-like 12 (proline rich)
bHLH	basic helix-loop-helix
CA11	carbonic anhydrase XI
CABP5	calcium binding protein 5
CAR	caspase-recruiting domain-containing protein
CD37	CD37 antigen
CGB2	chorionic gonadotrophin beta polypeptide 2
CGB7	chorionic gonadotrophin beta polypeptide 7
CMF	cardiac muscle factor
COL1A1	collagen type I, alpha 1
COL1A2	collagen type I, alpha 2
COL3A1	collagen type III, alpha 1
CRX	cone-rod homeobox
CSX	cardiac specific homeobox
Cx26	connexin 26
Cx37	connexin 37
Cx40	connexin 40
Cx43	connexin 43
Cx45	connexin 45
Cx46	connexin 46

Cx50	connexin 50
DHand	heart and hand neural derivatives, expressed 2
DFNA4	deafness, autosomal dominant nonsyndromic sensorineural 4
DMK	dystrophia myotonica protein kinase
dmk	mouse dystrophia myotonica protein kinase
EHand	heart and neural derivatives, expressed 1
ELSPBP1	epididymal sperm binding protein 1
EMP3	epithelial membrane protein 3
FCGRT	Fc fragment of Ig G receptor transporter alpha.
FLT3LG	fms-related tyrosine kinase 3 ligand
FTL	ferritin light polypeptide
FUT1	fucosyltransferase 1 (galactoside 2-alpha-L-fucosyltransferase Bombay phenotype included)
FUT2	fucosyltransferase 2 (secretor status included)
GATA 4	GATA 4 transcription factor
GPA	glycophorins A
GPB	glycophorins B
GRIN2D	glutamate receptor ionotropicN-methyl D-aspartate 2D
GRWD	glutamate rich WD repeat protein GRWD
GSK3A	glycogen kinase 3-alpha
GYS1	glycogen synthase 1
HLA	major histocompatibility complex
HRAS	Harvey rat sarcoma viral oncogene homolog
HRC	histidine-rich calcium binding protein
KCNA7	potassium voltage-gated channel shaker-related subfamily member 7
KIR2.4	potassium inwardly-rectifying channel, subfamily J, member 14
KDELR1	KDEL (Lys-Asp-Glu-Leu) endoplasmic reticulum protein retention receptor 1
KLK1	kallikrein 1
KPTN	actin-binding protein
LHB	luteinising hormone beta polypeptide

LIG1	ligase I DNA, ATP-dependent
LIN7B	lin-7 homolog B (*C. elegans*)
LDLR	low density lipoprotein receptor
LMNA	lamin A/C
MEF	myocyte enhancer factor
MSX2	Msh homeobox homolog 2
MTRNR1	ribosomal RNA, mitochondrial, 12S
MyBP	myosin-binding protein
MYH	myosin heavy chain
MYLIA	myosin light chain polypeptide 1A
NKX2-5	NK2 transcription factor related, locus 5
NOSIP	nitric oxide synthase interacting protein
NTF5	neurotrophin 5 (neurotrophin 4/5)
NTT5	neurotransmitter transporter 5
NUCB1	nucleobindin 1
PLA2G4C	phospholipase A2 group IVC (cytosolic calcium-independent)
PLEKHA4	pleckstrin homology domain containing family A (phosphoinositide binding specific) member 4
PPF1A3	protein tyrosine phosphatase, receptor type, of polypeptide, interacting protein (liprin)
PRPC8	precursor mRNA-processing factor 8
PSCD2	pleckstrin homology, sec7 and coiled-coil domain 2
RCV1	recoverin
RPL18	ribosomal protein L18
RPS11	ribosomal protein S11
RRAS	related RAS viral oncogene homolog
RUVBL2	ruvB-like 2 (*E. coli*)
SAP97	synapse-associated protein 97
SCN5A	sodium channel, voltage-gated, type V, alpha subunit

SLC8A2 solute carrier family 8 (sodium independent inorganic phosphate cotransporter), member 2

SLC17A7 solute carrier family 17 (sodium independent inorganic phosphate cotransporter), member 7

SNRP70 small nuclear ribonucleoprotein 70kDa polypeptide (RNP antigen)

SPHK2 sphingosine kinase 2 *DBP*D site of albumin promoter (albumin D-box) binding protein

SR-A1 serine arginine-rich pre-mRNA splicing factor SR-A1

SULT2A1 sulfotransferase family cytosolic 2 adehydroepiandrosterone (DHEA) – preferring member 1

SULT2B1 sulfotransferase family cytosolic 2B member 1

SYNGR4 synaptogyrin 4

TBC1D17 TBC1 domain family, member 17

Tbx2 murine T-box 2

TEAD2 TEA domain family member 2

TULP2 tubbylike protein 2

VRK3 vaccinia related kinase 3

ZNFP9 zinc finger protein 9

ZFP473 zinc finger protein 473 *(alias ZFP100)*

(http://www.ncbi.nlm.nih.gov/UNIGENE)

CHAPTER 1

INTRODUCTION

INDEX

1.1 ORGANISATION OF CHAPTER 1

This review forms the basis of my doctoral studies that involved the search for a mutated gene which causes a disorder named progressive familial heart block I (PFHBI:OMIM 113900) in a large South African family (Brink and Torrington, 1977). PFHBI is a cardiac disorder affecting the cardiac conduction system (CCS), specifically the bundle branches (BBs), and is detected by an electrocardiogram (ECG). To further an understanding of the aetiology of the disease, the structure, the development, ageing and electrophysiology of the CCS will be reviewed in the following section (section 1.2).

A number of other familial cardiac conduction diseases and syndromes are prevalent worldwide, for which in some cases the causative genes have been identified, namely, isolated cardiac conduction disease (ICCD) (Tan *et al.*, 2001), progressive cardiac conduction disease (PCCD) (Schott *et al.*, 1999) and idiopathic ventricular fibrillation (IVF) (Chen *et al.*, 1998). One could suggest that similar disease-causing mechanisms may be involved in the pathophysiology of PFHBI, it is therefore relevant to review the literature pertaining to some of these diseases. In addition, there are also syndromic diseases associated with cardiac conduction abnormalities; these include Holt-Oram syndrome (HOS) (Terret *et al.*, 1994), atrial septal defect (ASD) (Kahler *et al.*, 1966; Pease *et al.*, 1976; Basson *et al.*, 1995) and myotonic dystrophy (DM) (Harper, 1989) (section 1.3), which also render them important for discussion. Since PFHBI is the focus of this study, a detailed background and description of this condition is given of this disorder (section 1.3.2.3.4).

There are many different strategies available to search for a disease-causing gene and in the case of PFHBI, a positional candidate gene approach was adopted because the disease had been mapped to a specific locus and candidates were therefore selected based on position and cardiac function. The background for the strategies used in this study is reviewed in section 1.4. Previous linkage studies mapped the PFHBI locus to chromosome 19q13.3 (Brink *et al.*, 1994; Brink *et al.*, 1995), therefore, as part of this study, genes mapping to this region were investigated according to expression profile and function (see section 1.4.5). A list of candidate genes was prioritised for mutation screening in a PFHBI panel and the background and rationale for the selection of each candidate is discussed in section 1.5.

In most cases, the management of gene searches involves a group or a collaborative effort, as is the situation with PFHBI. Therefore, as examples of successfully managed disease gene identification strategies, the approaches followed for the genes causing retinitis pigmentosa (RP13) and keratolytic winter erythema (KWE) are discussed and compared to the PFHBI gene search (section 1.6).

1.2 CARDIAC CONDUCTION SYSTEM (CCS)

This section deals with the historical developments, the structure, the development, the ageing and the electrophysiology of the CCS.

1.2.1 Historical overview

Over the past number of years, there has been a paradigm shift from the previous understanding of the structure, development and functioning of the CCS (Davies *et al.*, 1983; Opie, 1991), which is partly due to research on *Mus (M) musculus*, *Gallus (G) gallus* and other vertebrates (Moorman *et al.*, 1998; Nguyen-Trán *et al.*, 2000; Pennisi *et al.*, 2002). The discovery of the CCS dates back to the mid-nineteeth century and investigations focused largely on the anatomy of the CCS (Davies, 1971; Luderitz, 1995). The evolutionary origin of the CCS has been a subject of much controversy, particularly as to whether the CCS develops from nervous or muscle tissue (see a more detailed discussion in section 1.2.3). As reviewed by Davies (1971), the myogenic theory was initiated by Gaskell in 1883, which proposes that the CCS is composed of specialised cardiac muscle cells; whereas the neurogenic theory, championed by Glomset and Glomset in 1940, proposed that the CCS develops from nervous tissue. Despite the overwhelming scientific evidence presented by Kistin (1949), Lev (1960) and Hudson (1965) that the CCS is myogenic in origin, electron microscopy studies in animal tissues show a close association between nerves and conduction fibres throughout the heart (reviewed by Davies, 1971). However, studies done by classic anatomists such as Purkinje, His, Tawara and Kent supported the myogenic theory (Davies, 1971).

The conduction of an electrical impulse in the heart is spontaneously generated by the specialised sino-atrial (S-A) node in the atrium, passed to the atrio-ventricular (A-V) node after spreading through the atria and ultimately to the His bundle, which delivers the

impulse to the BBs. It has been found that conduction through the His bundles and down the BBs to the ventricular endocardium is by Purkinje fibres (reviewed by Opie, 1991).

More recently, the different stages of cardiac development and the control of mRNA and protein expression profiles of the CCS have been investigated (Moorman *et al.*, 1998; Franco *et al.*, 1998; Gourdie *et al.*, 1999; Brunskill *et al.*, 2001). The data generated by these studies resulted in the construction of an integrated transcriptional and morphological model of the developing heart that will be discussed in section 1.2.3.

1.2.2 Structure of the adult ventricular cardiac conduction system

Since PFHBI is a disorder that affects the CCS, more specifically the conduction of the electrical impulse in the ventricles, a focus on the structure of the different components of the CCS, especially the ventricular and BB systems, is warranted. It can be proposed that any structural fault in the tissues comprising these systems could result in delay of the electrical impulse, thus causing PFHBI.

The conduction of the impulse, as described in section 1.2.1, results in the contraction of the heart. ECG studies of individuals suffering from PFHBI indicate a conduction block in the ventricular region, therefore, the anatomy of the A-V junctional and the A-V BB areas is discussed. These areas can be divided into five sections (Davies *et al.*, 1983; Brink, 1997) (fig. 1.1). They are (i) a transitional area composed initially of atrial working myocardium and then changing to the more compact cells of the A-V node, (ii) a compact area consisting of spindle-shaped cells similar to cells found in the SA node,

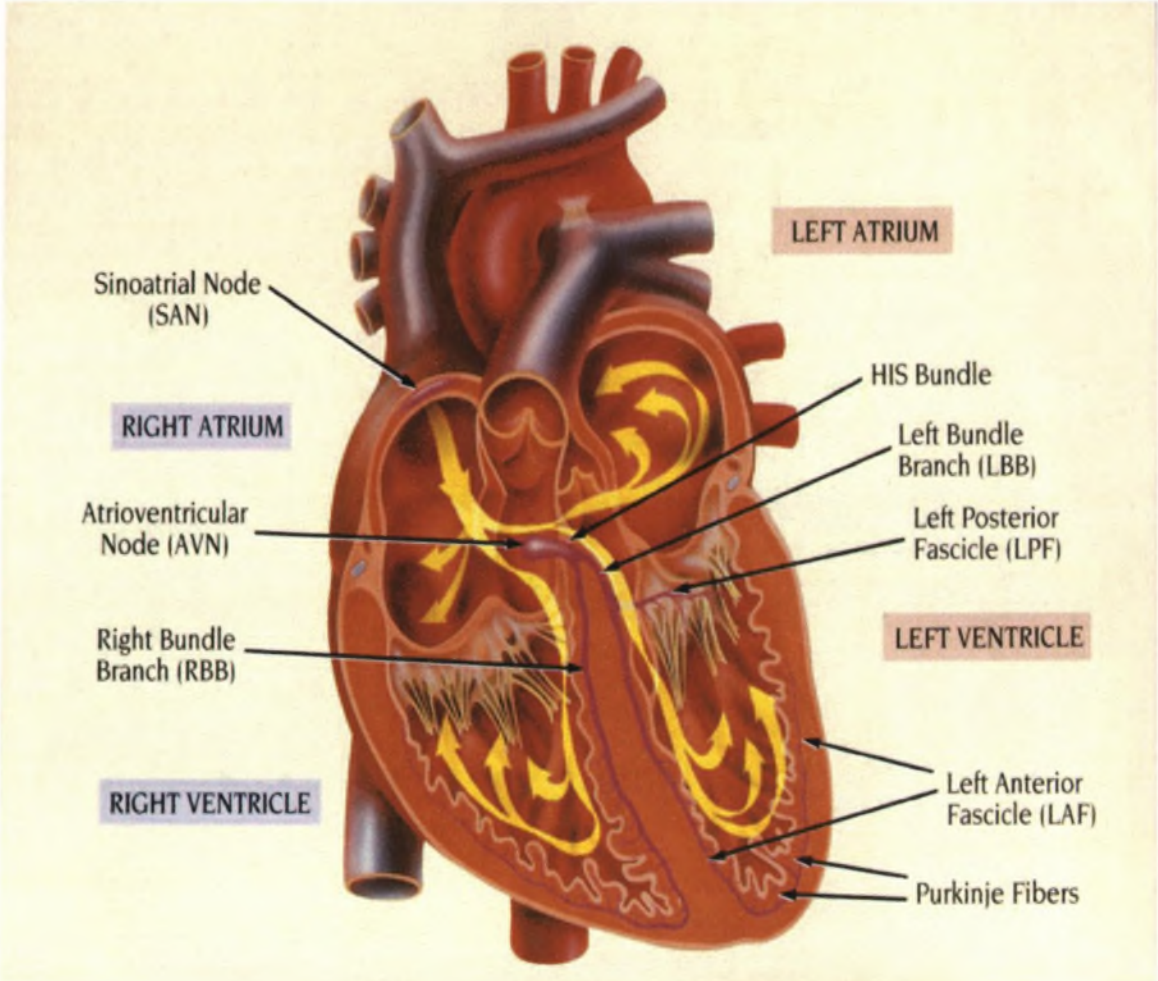


Figure 1.1 Different components of the cardiac conduction system (http://faculty-web.at.northwestern.edu/at/beaman/ecg/primer.html).

(iii) a transitional area with a presence of Purkinje type of cells found predominantly in the ventricular CCS, (iv) a non-branching ventricular component and (v) a branching ventricular component.

1.2.2.1 Atrioventricular junctional area and the atrioventricular bundle branch tissue

The transitional and the compact area in the atrial working myocardium, and the transition to the bundle of His, form the A-V node. The penetrating part and the non-branching ventricular part are referred to as the bundle of His (fig. 1.1) (Davies *et al.*, 1983). Like the S-A node, the A-V node depolarises spontaneously but is slow in its conduction properties, 5cm/sec, whereas the conduction velocity of the bundle of His bundle and its branches is 200cm/sec. The cells found in BB tissue are the Purkinje cells, which closely resemble embryonic muscle cells. The reason for the difference in the speed of the nodes and the BBs is possibly due to the difference in the anatomy between the nodes and the bundle of His. One explanation offered is that the nodes lack gap junctions and the bundle of His contains numerous gap junctions, which are presumed to be responsible for its fast conduction characteristics (reviewed by Brink, 1997). One of the major components of a gap junction is the protein connexin (Cx). A more detailed discussion of Cxs will be presented in section 1.2.2.3.

Another important feature of the A-V junctional area is that it is closely associated with the fibrous cardiac skeleton and therefore disease of the fibrous skeleton may have an effect on this area and may lead to A-V block (Davies *et al.*, 1983). On entering the ventricles and leaving behind the cardiac skeleton, the bundle of His branches into a left bundle (LBB) and a right bundle branch (RBB) system (Davies *et al.*, 1983). There are differences between the LBB and the RBB, for example, the RBB is a direct continuation of the main bundle of conduction tissue and can be seen by the naked eye as a single

distinct thin yellow streak due to the adipose tissue around the conduction tissue. It runs downward and forward into the septum towards the base of the ventricle. When it reaches the muscle, the right branch breaks up into a complex network of fibres, which run throughout the ventricle beneath the endocardium. In contrast, the LBB leaves the main BB as slender strands of fibres to form a wide fan of cells (Davies, 1971). The precise morphology of the left bundle has been a subject of considerable debate. For the purposes of this study, the focus is specifically on the RBBs since the disorder is characterised by RBB (Davies *et al.*, 1983). Histologically, the right bundle consists of parallel fibres, each 7.5μ in diameter, with few myofibrils and containing cells with a vacuolated cytoplasm (Davies, 1971).

Another explanation for the accelerated conduction velocity of the RBB is that it is due to conduction via the numerous end-to-end connections between Purkinje cells, rather than conduction via the side-to-side connections found in other myocardial cells (Opie, 1991). The cytoskeletal protein, desmin, has been found to be present in the Purkinje cells (Moorman *et al.*, 1998) and may be an important candidate gene for PFHBI. One could speculate that the cardiac conduction block found in PFHBI is caused by the malfunctioning of a defective (mutated) cytoskeletal protein. For this reason, any cytoskeletal gene present at the PFHBI locus should be considered an important candidate.

1.2.2.2 Proteins and ion channels of the conduction system

Guanine nucleotide-binding proteins (called GTPases or G proteins) are examples of a family of cytoplasmic signal-transducing proteins that are coupled to receptors, which are modulated by guanosine triphosphate (GTP). The binding of GTP to such a cellular receptor alters its conformation and increases its affinity to bind G proteins (Pollard and Earnshaw, 2002).

Eukaryotic cells use G proteins to regulate a host of functions ranging from transduction of signals from plasma membrane to regulation of the cytoskeleton, membrane trafficking, nuclear transport and protein synthesis (Pollard and Earnshaw, 2002). All these G proteins share a homologous core domain that binds a guanine nucleotide and use a common enzymatic cycle of GTP-binding, hydrolysis, and the product dissociation to switch the protein on or off.

The conformation of a G protein depends on whether GTP or GDP is bound. The GTP-bound conformation is the active form, as it interacts with, and stimulates, effector proteins and it is the γ–phosphate of the GTP, which turns the molecular switches on or off. The GDP-bound conformation is inactive because it does not bind the effectors. In many cases, binding to the effector proteins or regulatory proteins accelerates the activation step. Although there are five different families of G proteins, for the purposes of this review, only the trimeric G protein and the small G protein families will be discussed because of their roles in cardiac function.

Families of related genes in the mammalian cell encode the subunits of G proteins (Simon *et al.*, 1991).The number of genes in each mammalian gene family varies, e.g., α subunits are encoded by 16 genes, β subunits are encoded by five genes, and the γ subunits are encoded by 12 genes (reviewed by Downes and Gautam, 1999). The trimeric G proteins are heterotrimers made up of an α, β and γ subunits (reviewed by Downes and Gautam, 1999) (fig. 1.2). The seven-helix receptors activate trimeric G proteins by promoting the dissociation of GDP bound to an inactive trimeric protein. The GTP-binding changes the conformation of the G_{α} and dissociates the $G_{\beta\gamma}$. This generates two signals, as both G_{α} and $G_{\beta\gamma}$ can engage downstream effector proteins (fig. 1.2). There is evidence of at least 20 α, 6 β subunits and 12 γ subunits that comprise from 1-10% of the membrane protein of a cell and different combinations of these subunits can generate up to 1440 combinatorial signal transduction options (Zipes and Jalife, 2001).

In the heart, a G protein-based communication network is vital for relaying extracellular signals that mediate fundamental activities of the cardiac cycle such as electrical excitability, excitation-contraction coupling, calcium (Ca^{2+}) signaling, metabolic activity and phosphorylation events. The basic components of this system are the G protein-coupled receptor (GPCR), G protein complex of G_{α} and $G_{\gamma\delta}$ subunits and an effector (Zipes and Jalife, 2001). The two cardiac inward rectifier ion channel subunits, Kir 3.1 and Kir 3.4, and the voltage-gated fast acting sodium (Na^{+}) channel are under the direct regulation of G proteins (Zipes and Jalife, 2001).

The six families of small G proteins, such as the Rho and Ras families, consist of a single domain of 200 amino acid residues (fig. 1.2). Their functions include nuclear transport and regulation of the actin cytoskeleton. There are a number of proteins involved in the activation of the G proteins, for example, guanine nucleotide exchange factor (GEF), which stimulates the GTP hydrolysis that turns these molecules on or off (Hart *et al.*,

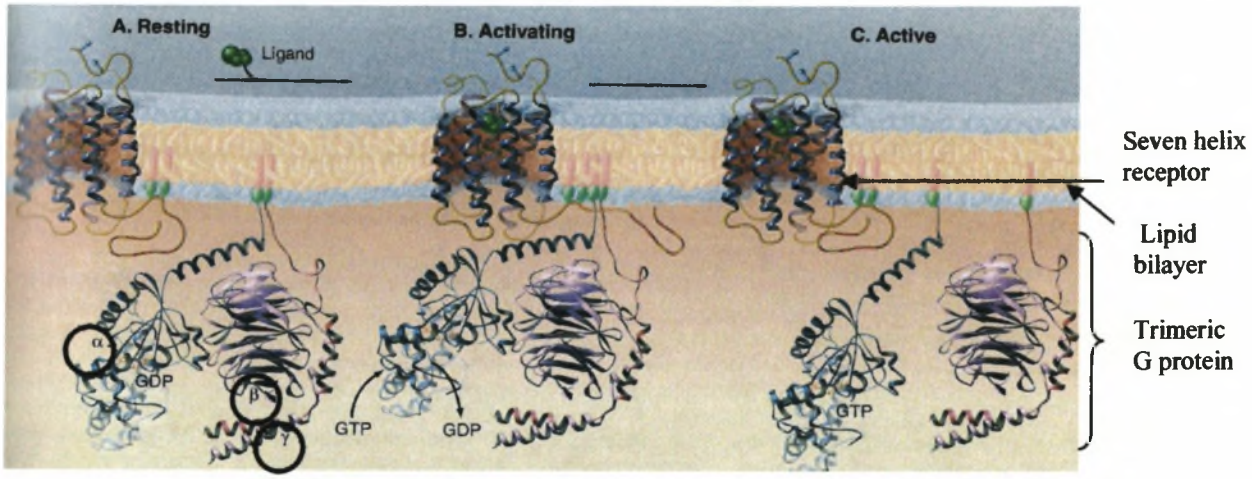


Figure 1.2 Atomic model and activation of a seven-helix receptor and trimeric G protein.
Subunits Gα, Gβ and Gγ of the trimeric G protein are circled. A, resting state, which lacks the binding of a ligand and the trimeric G-protein is in an inactive GDP-Gαβγ state. Both Gα and Gγ are anchored to the lipid bilayer. B, the binding of the ligand activates the receptor, which catalyses the exchange of GDP for GTP on Gα. C, active Gα and Gβγ dissociate from each other and the receptor become available to interact with effector proteins (Pollard and Earnshaw, 2002).

1996). Alterations in G protein-coupled receptors, or in the α subunits of the G protein, may lead to different human diseases with different phenotypes, such as pseudohypoparathyroidism type I, that is a disease caused by a defective α subunit of a G protein which results in an individual having shortened fingers, toes and stature (reviewed by Spiegel, 1996). McCune-Albright syndrome is also caused by a mutation in the α subunits of the G protein and is characterised by a classic triad of polyostotic fibrous dysplasia, café-au-lait skin hyperpigmentation and premature puberty (reviewed by Spiegel, 1996). For the most part, the diseases are confined to rare endocrine

disorders. A recent study suggests that mutations in G proteins can also lead to hypertension (reviewed by Farfel *et al.*, 1999). To date, a defective G protein has not been linked to any conduction, arrhythmia or primary cardiac disease.

In summary, G proteins are important in cardiac functions such as complex signaling pathways and ion channel function. One can, therefore, hypothesise that a mutation in any G protein or G-interacting protein encoding gene, could result in a cardiac conduction disorder such as PFHBI, and can be considered as plausible candidates for PFHBI.

Cardiac ion channels

A number of different cardiac ion channels controlling the flow of ions, such as the Na^+, Ca^{2+}, potassium (K^+) and chloride (Cl^-) ions, are present in the membrane of the myocytes of the heart. The movement of ions through these channels generates action potentials that ultimately activate the heart (Katz, 1993); Consequently, any change in gene expression or channel structure can be expected to result in a disruption of ion flow, which could lead to a cardiac conduction defect. One can, therefore, suggest that any ion channel gene, or any gene encoding a protein involved in ion channel function, would be a good PFHBI candidate. Ion channel genes and genes involved with ion channel function present at the PFHBI locus are discussed in more detail in sections 1.5.2 and 1.5.3.

1.2.2.3 Gap junctions

Gap junctions are specialised membranous intercellular channels, which connect adjacent cells in many tissues and organs, thereby providing chemical and electrical communication. In the heart, gap junctions provide pathways for intercellular current flow, enabling a coordinated propagation of action potential (reviewed by Bruzzone *et al.*, 1996; reviewed by Jalife, 1999; Jongsma and Wilders, 2000). Mammalian gap junction channels are built of Cxs encoded by a family of closely related genes.

Connexins (Cxs)

Four major Cxs have been shown to be associated with the myocytes, namely, Cx40 (α_5 connexin), Cx43 (α_1 connexin), Cx45 (α_6 connexin) and Cx37 (α_4 connexin) (Vozzi *et al.*, 1999; Zipes and Jalife, 2001). Different types of Cxs exhibit different channel properties and are regulated by specific gating mechanisms. The distribution and the relative abundance of Cxs in various tissues are different and studies with Cx37, Cx40, Cx43, Cx45 and Cx46 have shown that individual numbers and size of gap junctions differ in specific areas of the human heart and have characteristic conductive properties (reviewed by Bruzonne *et al.*, 1996). Research by Oosthoek *et al.*, (1993) revealed that Cx43 is not present in the myocytes of the A-V node, whereas abundant expression of Cx43 was found in the Purkinje fibres of the A-V bundle and the BBs. These differences are likely to play a role in regulating cardiac conduction properties and mutations in gap junction proteins may contribute to development of arrhythmias (reviewed by Jongsma and Wilders, 2000). It is hypothesised that this compartmentalised expression of Cxs is the way in which there is an orderly and sequential spread of activation from the atria to the ventricular chamber (reviewed by Gros *et al.*, 1996).

Each Cx is encoded by a single gene (*Cx*), which is composed of two exons separated by a large intron (fig. 1.3a). A typical Cx protein has four membrane-spanning regions and two extracellular loops and three cytoplasmic regions (fig. 1.3b). The original model of a complete intercellular channel was developed by Makowski *et al.*, (1977) and is still valid today. Six Cx proteins form a connexon, which spans the plasma membrane of two adjacent cells, so that each connexon contributes to one half of the channel. The two connexons interact in the extracellular space to form the complete gap junction channel, allowing direct communication between the cytoplasm of the participating cells (reviewed by Bruzzone *et al.*, 1996). A schematic view of the molecular steps leading to formation of an intercellular channel is shown in fig. 1.4. In the case of intercellular channels, there are no known substances that increase the open state, whereas closure of the channels can be observed in response to diverse treatment such as pH variation and changes in cytosolic Ca^{2+}. Studies have shown that Ca^{2+} levels of 400 to 500 nM can cause a channel to close. However, it is unlikely that there is a direct effect of Ca^{2+} on Cxs, since there are no Ca^{2+}-binding sites on Cxs. A recent hypothesis is that a Ca^{2+}-receptor protein, calmodulin, is the mediator of the Ca^{2+} effect. Another possibility is that the activation of specific kinases and subsequent phosphorylation of Cxs may mediate Ca^{2+}-gating (reviewed by Bruzonne *et al.*, 1996). It can be suggested that a defect in a Ca^{2+}-receptor protein, or a defect in a related kinase, may lead to an imbalance in the normal Ca^{2+} concentration, which could lead to cardiac conduction defects.

The use of transgenic knockout mice has advanced the understanding of the role of Cxs in cardiac development and function. A recently produced *Cx40* knockout mouse model

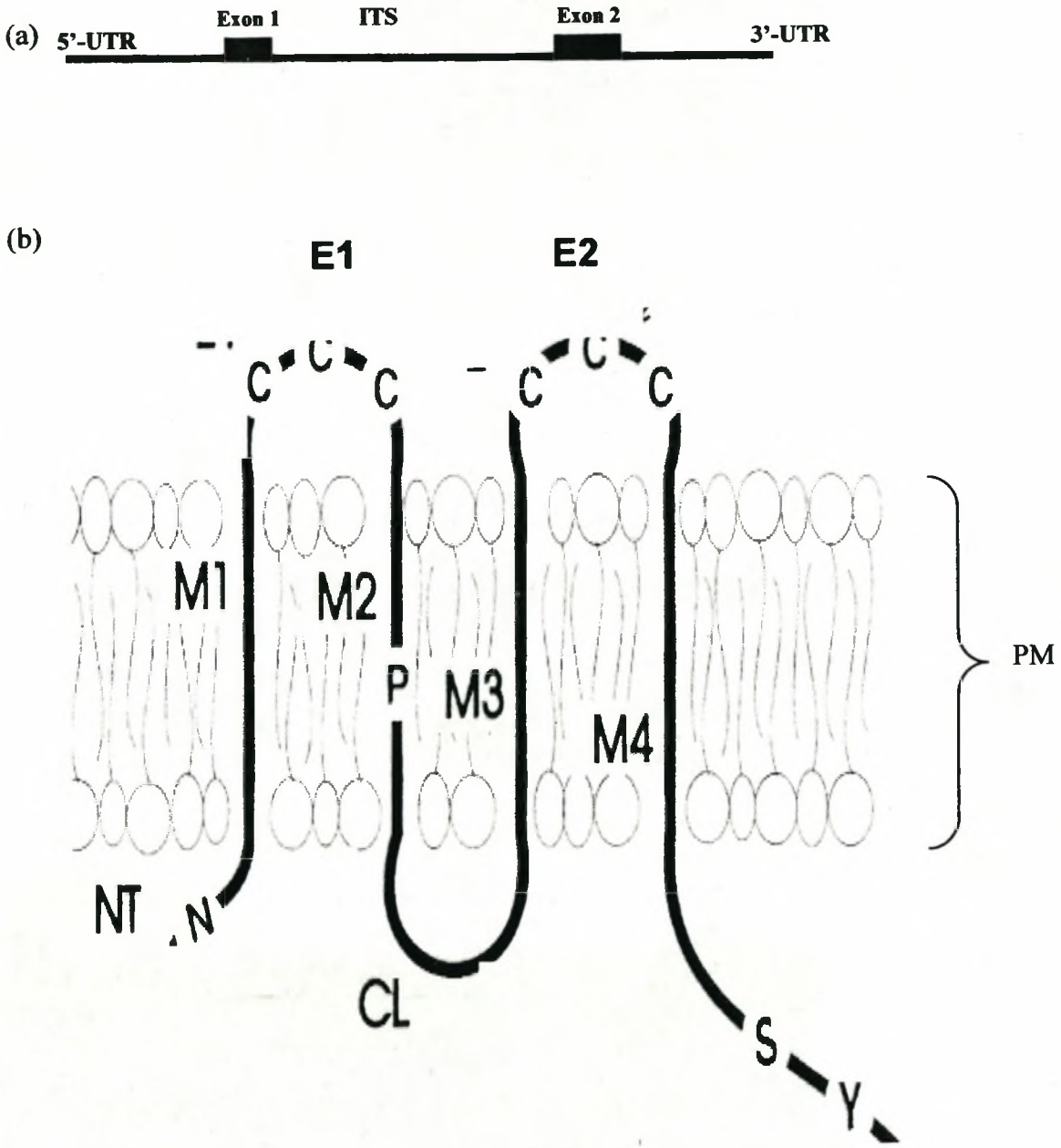


Figure 1.3 Schematic representation of a *Cx* gene and protein.
(a) Structure of a mammalian *Cx* gene. *Cx* genes consist of 2 exons separated by a large intron. 5'-UTR, 5'-untranslated region; 3'-UTR, 3'-untranslated region; ITS, intronic region.(b) Structure of a Cx protein. M1-M4, membrane-spanning regions; E1 and E2, the extracellular loops with spaced cysteine residues (C); NT, amino terminal; PM, plasma membrane; N, asparagine; P, proline; S, serine; Y, tyrosine (reproduced with minor modifications from Bruzzone *et al.*, 1996).

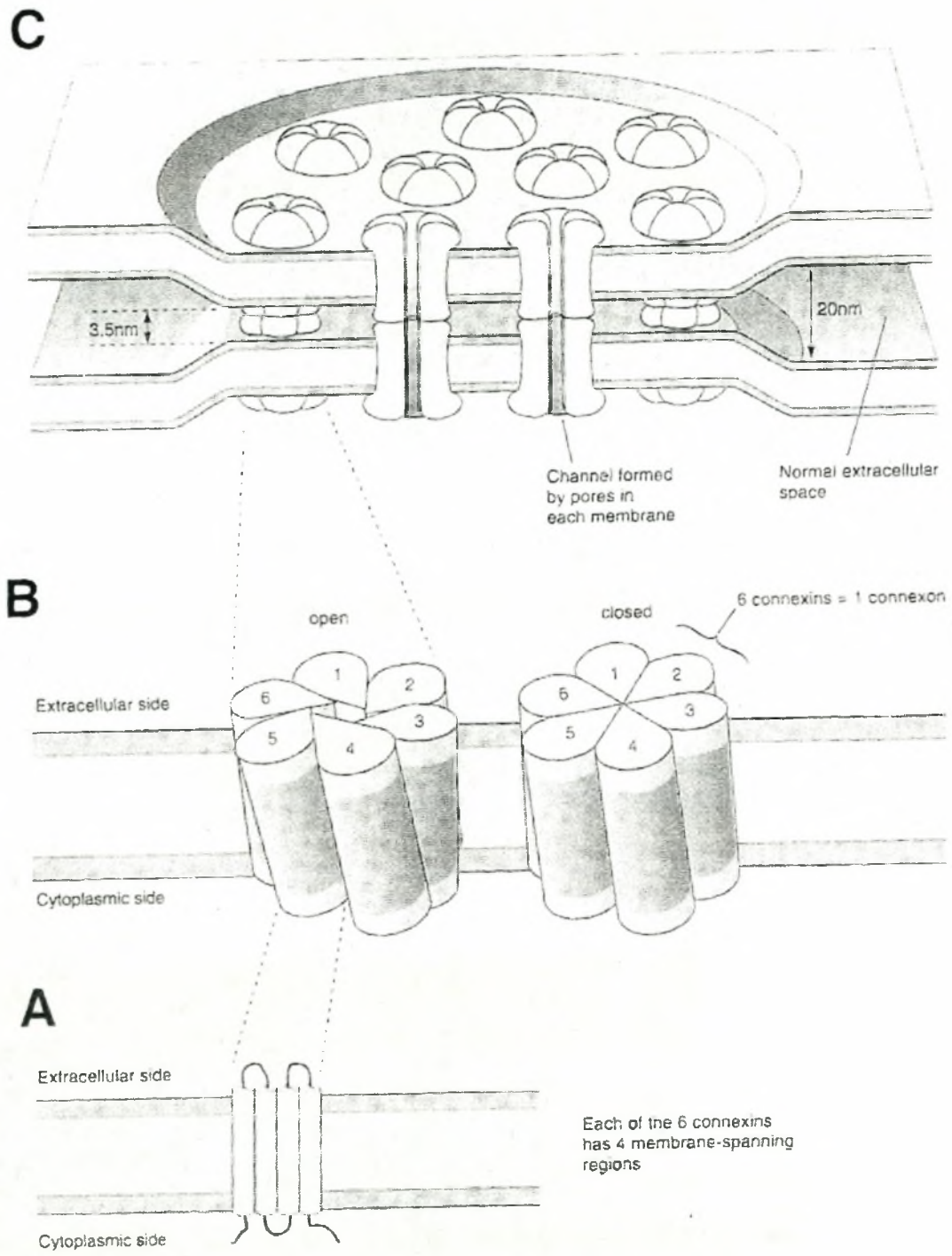


Figure 1.4 Schematic views of the molecular steps leading to the formation of intercellular channels. The protein subunits, the connexins (A), oligomerise in a hexameric structure, the connexons (B), that are transported to the plasma membrane. Connexons from adjacent cells interact to form complete intercellular channels (C) that become clustered in specialised membrane regions, the gap junctions (Bruzzone *et al.*, 1996).

showed cardiac conduction abnormalities characteristic of first degree A-V block and BBB (Simon *et al.*, 1998). Studies with a *Cx43/Cx40* double-knockout mouse model showed that Cx43 and Cx40 have additive effects on ventricular conduction and are also involved in heart morphogenesis (Kirchoff *et al.*, 2000). Electrocardiography revealed very small increases in the QRS and QT intervals recorded by ECG in the double heterozygous knockout mice ($Cx43^{+/-}/Cx40^{+/-}$), compared with the wild-type ($Cx43^{+/+}/Cx40^{+/+}$) or single heterozygous knockout animals ($Cx43^{+/-}/Cx40^{+/+}$;$Cx43^{+/+}/Cx40^{+/-}$). These changes were proposed to have arisen from the combined effects of slowed conduction in the central CCS (where *Cx40* is expressed) and in the working myocytes (where Cx43 is expressed). The functions of Cx45 were revealed by using *Cx45*-deficient mouse embryos, which exhibited striking abnormalities in vascular development. The embryos died by embryonic day 10 and exhibited a complex cardiac phenotype. In addition, there was a conduction block through the A-V canal and the contractions in the outflow tract were not co-ordinated (Kruger *et al.*, 2000). Analyses of mice with deletions of *Cx* genes have provided evidence that Cx43, Cx40 and Cx45, and, consequently, the gap junctional channels are involved in both heart function and development (Miquerol *et al.*, 2003).

Mutational alterations of *Cx* genes have led to several quite different diseases, indicating that these proteins have diverse functions. Individual mutations of *Cx40*, *Cx43,* and *Cx50* in *M. musculus* have been found to result in heart malformations or cardiac conduction defects (Jalife *et al.*, 1999; Kirchoff *et al.*, 2000). However in man, mutations in *Cx26*, can give rise to recessive nonsyndromic deafness or, more rarely, dominantly transmitted

deafness (Kelsell *et al.*, 1997). Recently, it has been shown that mutations at this locus are one of the most common causes of deafness. Mutations in *Cx30*, *Cx31* (reviewed by Krutovskikh and Yamasaki, 2000) and Cx43 (Liu *et al.*, 2000) also lead to deafness. To date, *Cx* mutations involving cardiac conduction abnormalities and severe heart malformations have only been observed in studies in mice lacking *Cx40* and *Cx43* genes. It is envisaged that findings and future studies will eventually lead to an understanding of the involvement of Cxs in heart morphogenesis and in cardiac conduction disorders of humans.

In summary, any mutation in proteins involved in the constitution of gap junctions, such as a Cx, which is responsible for the transfer of depolarising action potentials in the myocardium, are plausible candidate genes, which may be involved in the pathogenesis of PFHBI.

1.2.3 Development of the conduction system

The development of the cardiac conduction tissue of the heart is one of the most contentious topics that exists in the conduction field. The co-expression of neural and muscle genes has confounded the understanding of the origin of the CCS. However, it is believed that there are two possible origins, myogenic (Patten, 1956) or neurogenic (Gorza *et al.*, 1988; Vitadello *et al.*, 1990). Gorza *et al.*, (1988) presented evidence that the so-called "neural proteins" originated from a neural crest origin. Another school of thought strongly supports the development of the CCS, specifically the BBs and Purkinje fibres from a myogenic origin (Gourdie *et al.*, 1995; Cheng *et al.*, 1999). Present

knowledge of the development of the CCS is gained mostly from the study of animal embryonic material. However, it has been shown that there are a number of differences in cardiac development in different species (Davies *et al.*, 1983; Fishman and Chien, 1997).

For reasons discussed previously, the next section outlines the four stages of the development of the CCS, which are the foetal, the postnatal, the adult and the ageing, with more of a focus on the A-V and the BB regions.

1.2.3.1 Foetal, postnatal and adult development of the conduction system

Since 1970, histological differences between the developmental stages of the CCS have been studied (James, 1970; Davies *et al.*, 1983). It was shown that by the sixth to eighth week of foetal development, the human heart has attained most of its adult characteristics (James, 1970). At the initial stages of development, the small dark cells of the sinus are all similar but at a later stage some of these cells change and develop into the S-A node (James, 1970). It has been suggested that the S-A and the AV nodes develop from paired primordial sites, since they lie on either end of the atrium. These two nodes are the two most efficient pacemaking units for several reasons: firstly, an optimal route for the pacemaking signal to the two atria and ventricles is provided, secondly, an abundant adrenergic and cholinergic innervation is present and, thirdly, a large artery is organised around the sinus within a dense collagen framework (Davies, 1971). All the cells in the foetal S-A node are similar, while at birth two distinct types of cells are distinguishable, the P and T cells. The sinus nodal artery also appears during foetal development and eventually develops into the largest single artery. It is hypothesised that the pressure, due

to the blood flow in the artery during the development of the heart, has a significant effect on the S-A node's pacemaking rate (Davies, 1971). Recent studies using *G. gallus* have shown that coronary arterial beds are necessary for the recruiting of adjacent myocytes to differentiate into conduction cells (Pennisi *et al.*, 2002). At this stage, the foetal A-V node contains very little collagen; however, after birth, the collagen content of the S-A node increases rapidly until adult life. The collagen within the S-A node is intricately woven among the P and T cells to form a regular pattern framework, whereas the internodal and interatrial pathways consist of largely fat and collagen and become fully defined during the postnatal period (Davies, 1971).

One could hypothesise that any structural gene expressed in the CCS, such as collagen, would be a plausible PFHBI-causing gene, since the increase in the collagen content during development coincides with the progressive course of this disease. This theory was further supported when an earlier weakly positive linkage result suggested that the PFHBI locus might map to chromosome 4, to a locus which encompassed three plausible collagen genes, collagen type I, alpha 1 (*COL1A1)*, collagen type I, alpha 2 *(COL1A2)* and collagen type III, alpha 1 (*COL3A1*). Subsequently, these genes were excluded as part of the same study (Brink *et al.*, 1994).

Moorman *et al.*, (1998) described the structure of the CCS of the embryonic heart as a cardiac tube with polarity. The atrial end of the tube contains the fast-conducting cells and the other end, the ventricular end, contains the slower conducting cells, so that the tube has a slow and fast area. This would explain how the electrical architecture of the embryonic heart is the same as the adult heart (Moorman *et al.*, 1998).

In 1970, James presented two theories about the embryological origin of the His bundle with its proximal branches. One theory states that the His bundle grows forward from the original A-V node and the other suggests that the His bundle and its proximal branches originate *in situ* within a primitive ventricular tissue and simply join the A-V node. In support of these theories, it has been observed that shreds of A-V nodal tissue are strewn along margins of the central fibrous body (including the crest of the ventricular septum) in the human infant heart. The tissue gradually becomes absorbed during postnatal development so that it is rarely present in the adult heart (Davies, 1971). Recently, it has been reported that the fast conducting tissue of the His-Purkinje system from human embryonic hearts of 4.5 to 10 weeks showed that the A-V node develops from the canal myocardium and retains its continuity with the ventricular myocardium (Kim *et al.*, 2001).

The His bundles are composed exclusively of typical Purkinje cells, as are the BBs. Gourdie *et al.*, (1995) provided evidence that the lineage of the Purkinje fibres develop from a myogenic origin. One set of experimental data suggests that the Purkinje fibres and working myocytes share a common myogenic precursor in the embryonic tubular heart (Gourdie *et al.*, 1995). Another set of data provides evidence that the differentiation of Purkinje fibres may occur before the differentiation of working myocytes. A third set of data suggests that the peripheral (i.e., the intramural Purkinje fibre network) and the central components of the CCS are derived from independent parental cells, and that two components are linked together to establish the sequentially integrated CCS (Gourdie *et al.*, 1995).

Other evidence suggests that the His bundle and its branches originate separately. If this is the case, then the joining of the His bundle and its branches occurs during early foetal development since they are united during the second month of gestation (Davies, 1971). The mechanisms governing the differentiation of these cardiac tissues are still poorly understood (Gourdie *et al.*, 1995). The cell types at the junctions are predominantly T cells at the A-V node and Purkinje cells in the His bundle. At the border region, a mixture of the two types of cells is found (Davies, 1971). More recently, it was suggested that the distal CCS, such as the A-V node, His bundle and BB, is derived from working mocytes (reviewed by Pennisi *et al.*, 2002). This was demonstrated in a series of retroviral lineage studies in chick embryos (reviewed by Mikawa, 1999).

To achieve its adult form, the His bundle undergoes extensive postnatal remodeling, which includes apoptosis (programmed cell death), just as morphogenesis often does (Davies, 1971). Apoptosis is a tightly regulated series of energy-dependent molecular and biochemical events orchestrated by a genetic programme (Colucci, 1996). Alterations in the apoptotic pathways can result in a number of diseases in humans, for example, apoptosis occurs in the myocardium of patients with arrhythmogenic right ventricular dysplasia, indicating that apoptosis is found in diseased hearts (Narula *et al.*, 1996).

The volume of collagen in the different tissues of the adult heart is so large that it is the most readily distinguished protein on staining. It has been found that diseases in which the collagen framework is damaged are associated with arrhythmia because the S-A node becomes unstable (James, 1971). In the adult CCS, three pathways connect the S-A node

to the A-V node and a fourth one connects the S-A node to the left atrium (fig. 1.1). The pathways are not ensheathed but are distinguished anatomically by the presence of the myocardial cells, which consist of a mixture of Purkinje and ordinary working myocardial cells. The conduction speed is more rapid than in the working myocardium, although not as rapid as in the His bundle and its branches (Pennisi *et al.*, 2002).

During the development of the CCS from the foetal to the adult form a number of processes, such as the differentiation of cells, an increase in collagen content of the CCS and extensive remodeling, occur. It can be suggested that any imbalances in one of these processes may lead to PFHBI.

Molecular and genetic developments

It was proposed by Moorman *et al.*, (1998) that, if the development of the CCS were to be described at the molecular level, the following events would be relevant: (i) the development of the cardiac tube, (ii) the development of the polarity of nodes, (iii) the chamber formation, (iv) the development of an electric impulse and (v) the involvement of Cxs, such as Cx37, Cx40, Cx43, Cx45 and Cx46; contractile proteins, such as α and β myosin light chain (MLC); cytoskeletal proteins, such as desmin, neurofilament proteins and other proteins such as, atrial natriuretic factor (ANF), acetylcholinesterase and creatine kinase.

In addition, a review by Franco *et al.*, (1998) of recent studies that investigated the genetic mechanisms controlling cardiac development has given good insights into the

development of the CCS. This review covers the patterns of expression of a number of proteins that have been studied previously at the three developmental stages (tubular, embryonic and foetal heart) in *Homo (H) sapiens*, *M. musculus*, *Rattus norvegicus* and *G. gallus*. The proteins investigated were (i) the contractile proteins, such as MLC, actins, tropomodulin, tropomyosin, troponins, calponin and titin; (ii) Ca^{2+} handling proteins, such as the sarcoplasmic reticulum Ca^{2+} ATPase, phospholamban, ryanodine receptor, Na^{+}-Ca^{2+} exchanger and Na^{+} K^{+} ATPase; (iii) impulse conduction proteins, such as Cxs; (iv) energy metabolism proteins, such as creatine kinase; (v) intermediate filament proteins, such as desmin; (vi) cell-cell and cell-matrix molecules, such as integrins and cadherins; (vii) cardiac specific transcription factors, such as the NK homeobox genes, GATA-binding proteins (GATA), myocyte enhancer factor genes (*MEF)* and basic helix-loop-helix genes (*bHLH*); (viii) non-cardiac specific transcription factors, such as serum response factor, caspase-recruiting domain-containing protein (*CAR)* and cardiac muscle factor (*CMF)* and (ix) other proteins, such as the *ANF* and acetylcholine esterase (Franco *et al.*, 1998).

Using all the data from these studies, a transcriptionally regulated model of the developing heart was proposed. For example, the development of the straight cardiac tube involved the expression of transcription factors *MEF-2*, heart and neural crest derivatives expressed 2 *(dHAND)* and GATA 4 (*GATA-4)*, in opposite gradients to each other. In the case of the looped heart, five functionally different segments can be distinguished, i.e., the inflow tract, the atria, the A-V canal, the ventricles and the outflow tract. During this stage, the expression patterns of a number of genes are either changed, remain the same

or are switched off which implies that a negative domain-specific transcription factor (repressor) is present. This hypothesis is supported by the change of expression patterns of several transcription factors, such as *eHAND* and *dHAND* and *GATA-5* (Franco *et al.*, 1998). Interestingly, on examination of the heart of a human embryo that had died of cardiac arrest, it was found to have no right ventricular chamber and it was also discovered that the transcription factor gene *dHAND* was deleted (Srivastava, 1997; Barinaga, 1998). In the third stage of development, the heart becomes septated and distinct right and left atrial and ventricular components are recognisable. A possible model of right-left transcriptional specification has been suggested, similar to the segmentation of the fruit fly, *Drosophila (D) melanogaster,* but the experimental evidence for this model has not been finalised (Franco *et al.*, 1998).

Another study that attempted to unravel the cell fate, lineage and morphogenetic decision that configure the vertebrate heart form and function in *D. melanogaster* and *Caenorhabditis (C) elegans* came to the conclusion that genetic analysis of these two organisms does not give all the answers. In a review by Pennisi *et al.*, (2002) it was suggested that further studies using *M. musculus* and *Danio (D) rerio* may provide more answers about cell fate, lineage and morphogenic processes.

Transcription factors have also been implicated in the regulation of heart development and conduction defects, e.g., genetic studies have shown that cardiac conduction defects in mouse models are linked to a mutation in a homeoprotein, namely, NK2 transcription factor related, locus 5 (*NKX2-5).* Habets *et al.*, (2002) found that *ANF* is expressed in the

developing chamber, but is absent in the nodal tissue from fish to humans. The repressor gene, *Tbx2,* is expressed in the embryonic heart in a mutually exclusive pattern to *ANF* which led to the proposal that Tbx2 protein could be halting the expression of the *ANF* gene in certain tissues such that they acquire nodal characteristics, rather than developing into chambers. They also discovered that Tbx2 binds transcription factor NKX2-5 and that the complex blocks the expression of the promoter region of the *ANF* gene in tissues flanking the chambers.

At the 2003 meeting of the Novartis Foundation Symposium, the traditional hypothesis of ion pumps, channels and Cxs (ICC)-centric model, which suggests that the abnormal functioning of ion pumps, channels and Cxs causes conduction defects, was discussed (Miquerol *et al.*, 2003). It was concluded that the ICC model had shortcomings and therefore two additional hypotheses were presented for the pathogenesis of conduction defects and arrhythmias. The one hypothesis proposed that conduction defects may arise from anatomic underdevelopment of the CCS *in utero* and the other hypothesis proposed that cardiac arrhythmias associated with *Nkx2-5* mutations may result from the non-uniform alteration of cardiac myocyte channel proteins, leading to increased electrical heterogeneity. In addition to the traditional ICC-centric model, these two alternative hypotheses are under investigation, which should lead to a richer understanding of cardiac conduction defects.

It is proposed that an understanding of the role and interplay of transcription factors, ion

channels and *Cxs* during the development of the CCS will give a clearer idea of how mutations in one of the genes encoding a cardiac conduction protein can lead to a cardiac conduction defect, such as PFHBI.

Animal studies

Studies of the CCS using *D. rerio*, *M. musculus* and *C. elegans* have yielded interesting findings and insights into possible causes of conduction defects (Stainier *et al.*, 1996; Gourdie *et al.*, 1999). These organisms are useful models for this type of study, because they reproduce and develop rapidly. A vascular cytokine, endothelin (ET), was shown to induce avian embryonic cardiomyocytes to differentiate into Purkinje fibres but as the embryo matured the ability to convert to Purkinje fibres decreased (Kanzawa *et al.*, 2002). Additional studies, using the avian system, showed that when muscle cells converted to Purkinje fibres, they expressed a unique myogenic transcription programme before and after the conversion (Takebayashi-Suzukie *et al.*, 2001).

D. rerio is an excellent model system in which to study cardiovascular development because a beating heart, which is functional within one day of fertilisation, is formed. The zebrafish embryo does not require blood flow for survival during the first two days of development (Dempsey *et al.*, 2000). Embryo research on *D. rerio* has given insight into the functioning of the pacemaker and the conducting system, because the embryo is transparent, the detection of rhythm is easily visible with the naked eye (reviewed by Chen and Fishman, 2000). A number of mutations in different genes, that have been imaginatively named, have caused a number of different cardiac conduction defects such

as, a slow heart beat, *slo mo* mutant, interference with conduction between chambers, *breakdance* and *hip hop* mutants, or a disturbance of conduction between cells, *island beat*, *polka* and the *tremblor* mutants (Stainier *et al.*, 1996). The identification of these novel mutants present in *D. rerio* may help to identify human gene homologs with similar functions that could possibly aid in the identification of the PFHBI-causative gene.

1.2.3.2 Ageing of the conduction system

Except for histological studies described by Davies and Pomerance (1972), very little research has been done in the past on the ageing process of the CCS. Investigations have shown that, with ageing, the amount of fibrous tissue in the CCS increases (Mandel, 1980). Fibrous tissue becomes more evident as the content of nodal tissue per unit area of node decreases with increasing age (Mandel, 1980). Additionally, in the ventricular CCS of healthy individuals, an increase in fibrosis present per unit area increases with age. Processes causing this increase are not fully understood. In conduction disease, a replacement fibrosis of the bundle of His and the BBs features prominently. This phenomenon is called "idiopathic BB fibrosis,'' as it appears that the BB tissue is dying off (Davies *et al.*, 1983). In 1970, Lev described a condition where adjacent myocardium is normal, but "sclerosis of the left-sided CCS" was present. Lenègre (1964) described a similar condition, but in this case the RBB was affected. Davies *et al.*, (1983) suggested that the conditions described by Lenègre and Lev were the same disease.

Recent progress in the understanding of the genetic controls involving ageing is driven largely by the use of model systems, ranging from yeast and nematodes to mice. It

appears that maintaining the genome integrity is a major factor in reducing ageing (Hasty *et al.*, 2003). If this is the case, one can also propose that an increase in the fibrosis of the conduction tissue in cardiac conduction disease is probably due an imbalance in a genetic pathway, which initially maintained the conduction tissue.

1.2.4 Electrophysiology

From the 1950s to the late 1970s, the explanation of the spread of currents in the cardiac muscle was based on the "continuous conduction theory". Since the discovery of the presence of gap junctions between cardiac cells, more experiments were performed and electrophysiological data were re-analysed; thus the concept of discontinuous conduction has gained more agreement (Spach, 2003).

A number of electrophysiological events lead to the excitation-contraction coupling in myocardial cells. Figure 1.5 shows the pattern of conduction from the SA node to the ventricles. The SA node has three depolarising currents, of which the slow upstroke is largely Ca^{2+}-dependent. In other tissues, such as the atrial muscle, Purkinje fibres and the ventricular muscle, the impulse travels very rapidly because the initial phase of depolarisation is much faster and is Na^{+}-dependent. The Ca^{2+}-dependent A-V node slows down conduction between the atria and ventricles.

1.2.4.1 Depolarisation

In the resting state, myocardial cells are polarised, meaning that they are carrying an

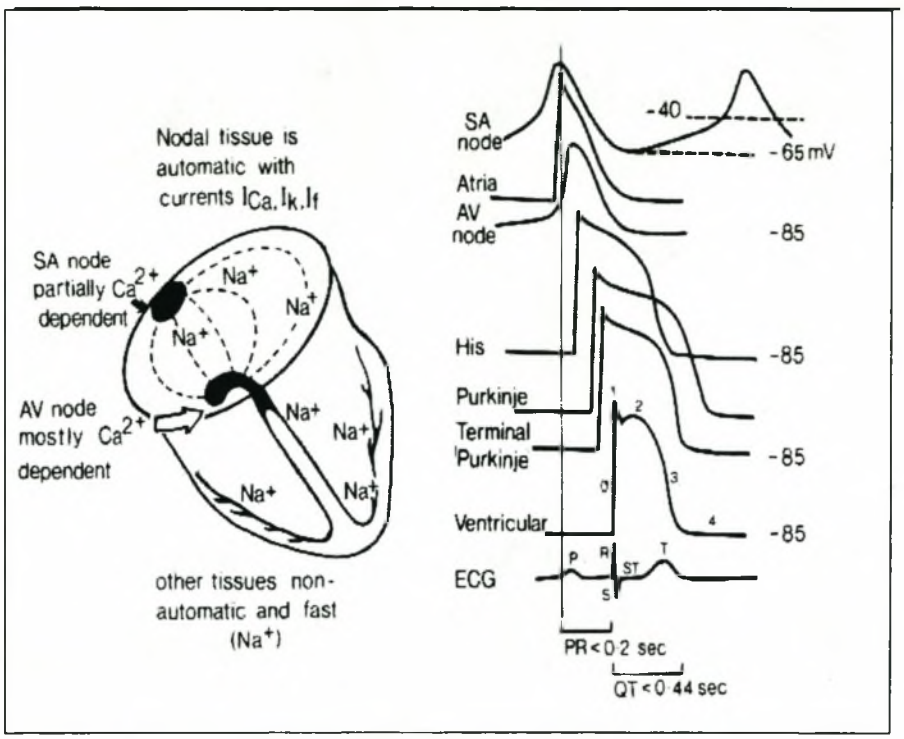


Figure 1.5 Pattern of conduction from the SA node to the ventricles.
On the left is a schematic drawing of components of the cardiac conduction system, indicating the depolarising ionic currents in the nodes. On the right are traces showing the action potential generated by the different ion changes in the conduction system and the ventricles (Opie, 1991).

electrical charge. These cells are positively charged on the outside and negatively charged on the inside and the potential difference across the membrane is approximately -90mV. When an electrical current is transmitted, or the cell is stimulated, it begins to depolarise at –65mV. At -40mV the activation threshold is reached. At this point the voltage dependent ion channels facilitate the movement of Na^+ and Ca^{2+} ions into the cell (Davies *et al.*, 1971). The inward flux of Na^+ and Ca^{2+} into the cytosol of the cell generates an action potential, which results in the contraction of the myocyte. The activation of Na^+ channels is much faster than that of the Ca^{2+} channels (Katz, 1993).

Consequently, the conduction of the electrical impulse in the rapidly conducting ventricular CCS is dependent on the Na^+ channel-generated action potential, with the Ca^{2+} channel generating secondary inward currents (Opie, 1991; Katz, 1993). In contrast, the slow conducting nodes depend mostly on the slow Ca^{2+} ion channel-generated action potential. In the A-V node, the conduction delay is largely due to the slow conduction rate mediated by the Ca^{2+} channels. The same depolarising stimulus responsible for the activation of the ion channels effects their inactivation, thus closing the Na^+ and the Ca^{2+} channels (Opie, 1991; Katz, 1993).

1.2.4.2 Repolarisation

The return of the cells to their resting state is affected by the outward movement of K^+ through K^+ channels, possibly supplemented by the inward movement of Cl^-. The two K^+ ion currents involved in this process are the transient outward current (I_{to}), which is responsible for the early repolarisation and the delayed rectifier (I_k), which is responsible for the delayed repolarisation and also for maintaining repolarisation (Katz, 1993).

The electrocardiogram

The ECG patterns of individuals suffering from PFHBI are distinguished by the presence of a broad QRS complex (Brink and Torrington, 1977), (fig. 1.8) which is caused by RBBB, and results in the late depolarisation of the ventricles (Hampton, 1997). The ECG is a non-invasive technique that records the cumulative currents produced by the heart muscles, thereby providing valuable information about the heart's normal or abnormal function (Goldberger and Goldberger, 1990). The ECG records the effect of the

depolarising stimulus on the myocardium (i.e., the sequential contraction of the atria and then the ventricles) as deflections, the P wave and the QRS complexes. The return to the resting state, repolarisation, is represented by the ST-segment and the T wave (fig. 1.5). In a healthy heart, the P wave represents the contraction of both atria in response to the depolarising stimulus initiated at the S-A node. The P wave is followed by a flat baseline, and represents the delay of the electrical impulse at the A-V node. Contraction of both ventricles is represented as a QRS complex. Following the QRS complex is a gently upsloping ST segment which precedes a T wave depicting ventricular re-polarisation. One cardiac cycle is represented by the P wave, QRS complex and a T wave.

1. 3 CARDIAC CONDUCTION DISEASES

The CCS is responsible for the conduction of electrical impulses in the heart and therefore the impairment of any one of the components of the CCS can result in heart rhythm disturbances (arrhythmias). These arrhythmias may progress to heart block, which may be fatal if not treated by timely implantation of an artificial pacemaker (Van der Merwe *et al.*, 1988). A number of cardiac conduction disorders have an inherited congenital basis implying that they include a genetic component and can be inherited in a family setting (Vallianos and Sideris, 1974; Esscher *et al.*, 1975; James *et al.*, 1975; Brink and Torrington, 1977; Brink *et al.*, 1995; De Meeus *et al.*, 1995). The generation of congenital cardiac defects is thought most likely to occur during the complex morphogenesis of the heart.

1.3.1 Familial congenital cardiac diseases

Cardiac development is a complex biological process requiring the integration of a variety of processes involving a number of cell components, morphogenesis and excitation-contraction coupling (Olson and Srivastava, 1996). A number of postnatal morphogenetic events only occur several months after birth (James and Facc, 1985). An understanding of these finely controlled molecular mechanisms has come from studies of *D. melanogaster*, *G. gallus* and *M. musculus* (Levin *et al.*, 1995; Rossant, 1995; Park *et al.*, 1996). Studies using *M. musculus* are laying the groundwork for a comprehensive understanding of cardiac organogenesis. One study involving *M. musculus* illustrated that the two related transcription factors, dHAND and eHAND, are expressed in a complementary fashion in the development of the right and left ventricles, respectively. Targeted mutations of these transcription factors revealed novel pathways of the development of the heart (Srivatava, 1997). The basis of congenital heart defects is still obscure, but disorders such as Holt-Oram syndrome (HOS) and atrial septal defect (ASD) have implicated mutations in transcription factor genes as being involved in heart development (Basson *et al.*, 1997; Schott *et al.*, 1998).

1.3.1.1 Holt-Oram sydrome (HOS)

Holt-Oram syndrome (HOS:OMIM 142900) is an autosomal dominantly inherited disorder that is characterised by skeletal abnormalities and cardiac malformations, such as either ASD, muscular ventricular septal defects (VSD) or multiple complex malformations (Holt and Oram, 1960; Basson *et al.*, 1994). From classical genetic linkage studies, the *HOS* gene was mapped to chromosome 12q21.3-q22 (Terret *et al.*,

1994). Using exon trap analysis of genomic clones from this interval, two developmentally expressed genes (T-box-3, *TBX3*, and T-box-5, *TBX5*) from the *Brachyury* (T) gene family were identified (Li *et al.*, 1997). Basson *et al.*, (1997) refined the mapping of the *HOS* locus to chromosome 12q24.1 by fluorescence in situ hybridisation (FISH) using a cosmid containing *D12S129*, which was tightly linked to the *HOS* gene. From the critical region, Basson and colleagues (1999) isolated a gene with a high degree of homology to *M. musculus Tbx5* and identified several mutations in *TBX5* in affected members of the HOS families. To date, ten different mutations have been found in *TBX5* in different affected families. Members of the *TBX* gene family act as transcription factors and the conserved TBX domain serves as a DNA-binding domain (Basson *et al.*, 1999). Hiroi *et al.*, (2001) found that TBX5 associates with NKX2-5 and synergistically promotes cardiomyocyte differentiation. Additional TBX5 protein-binding site studies have shown that the presence of *TBX5*-binding sites in the upstream regions of several cardiac-expressed genes, including cardiac α-actin, *ANF*, α-cardiac *MYH*, β-cardiac *MYH*, *MYL1A*, *MYL1V* and *NKX2-5,* suggests a role for TBX5 in their regulation (Ghosh *et al.*, 2001).

1.3.1.2 Atrial septal defect (ASD)

Atrial septal defect (ASD:OMIM 108900) is an autosomal dominantly inherited disorder that is characterised by cardiac malformations and A-V conduction abnormalities. Kahler *et al.*, (1966) reported three families with this combination and Pease *et al.*, (1976) and Basson *et al.*, (1995) identified four more families with ASD conduction defects only. Shiojima *et al.*, (1995) mapped the cardiac specific homeobox gene (*CSX*) to

chromosome 5q34, using FISH and systematic screening of a yeast artificial chromosome (YAC) library using polymerase chain reaction (PCR). *CSX* is expressed in various tissues including the CCS of the developing heart. Schott *et al.*, (1998) showed that ASD-affected individuals had mutations in the *CSX* gene, which encodes the transcription factor NKX2-5. In addition, another homeobox-containing gene, Msh homeobox homolog 2 (*MSX2),* was also mapped to this region. Shiojima *et al.*, (1995) suggested that localisation of *CSX* and *MSX2* to the same region of the genome may indicate that they are coordinately regulated during human heart formation.

In studies by Shiojima *et al.*, (1995), three different *NKX2-5* mutations were identified, of which two are predicted to impair binding of NKX2-5 to target DNA, resulting in haploinsufficiency, and the third potentially aids target-DNA binding. These investigations indicate that NKX2-5 is important for regulation of septation during cardiac morphogenesis and for maturation and maintenance of A-V node function throughout life.

1.3.2 Familial cardiac conduction diseases

Various forms of familial conduction diseases have been reported throughout the world; pedigrees of six South African families with hereditary A-V block have been reported (Myburgh and Steenkamp, 1973). In an investigation in the United States of America (USA), 26 of 59 members in four generations of a family showed abnormal ECGs. The affected members had conduction abnormalities as the only manifestation of the disease, inherited in an autosomal dominant manner (Schaal *et al.*, 1973). In another study, five

out of 14 family members presented with conduction defects, such as A-V block, RBBB, and LAHB. Another sibling died suddenly, six did not have conduction abnormalities and information was not available regarding the remaining two (Vallianos and Sideris, 1974).

There are also a number of inherited neurological disorders associated with cardiac arrhythmias that occur with either a direct involvement of the heart or induced neurohormonal abnormalities that act on the heart (Zipes and Jalife, 2000). Myotonic dystrophy (DM) and limb-girdle muscular dystrophy type 1B (LGMD1B) are examples of diseases that show both neurological symptoms and cardiac abnormalities, including A-V conduction disturbances, RBBB, LBBB and dysrhythmias, and in some individuals this necessitates pacemaker implantation.

1.3.2.1 Myotonic dystrophy (DM)

Myotonic muscular dystrophy 1, or Steinert's disease (DM1:OMIM 160900), is a disease showing an autosomal dominant form of inheritance. It is characterised by weakness and atrophy of distal skeletal muscles, as well as systemic manifestations, such as early balding, gonadal atrophy, cataracts, mental retardation and cardiac involvement (Harper, 1989; Perloff *et al.*, 1984). The myotonic dystrophy kinase, *DMK*, encodes a protein homologous to a serine/threonine protein kinase (Aslanidis *et al.*, 1992; Buxton *et al.*, 1992). DM1-affected families illustrate the phenomenon of anticipation (see section 1.3.2.3.4 for more detail on anticipation), which shows an increase in the number of CTG triplet repeats in the 3'-untranslated region (3'-UTR) of *DMK* gene (Brook *et al.*, 1992; Mahadevan *et al.*, 1992). Studies using DM1-affected individuals have shown that the

cardiac conduction abnormality increases both with age and with the extent of CTG repeat expansion (Melacini *et al.*, 1995; Wang *et al.*, 1995; Clarke *et al.*, 2001). Since the phenomenon of anticipation has also been observed with PFHBI (Van der Merwe *et al.*, 1988), one can speculate that PFHBI might be caused by a similar mechanism to DM1, implying an expansion of a triplet repeat in the PFHBI disease-causing gene.

The mechanism by which the amplified CTG repeat expansion leads to DM1 is unclear, however, a number of theories have been suggested, for example, an alteration in tissue phosphorylation due to a decrease in DMK activity which, in turn, has been postulated to cause a change in ion channel structure or function (Rudel *et al.*, 1989; Timchenko *et al.*, 1995). Mounsey *et al.*, (2000) used *M. musculus* myocytes that were either heterozygous (+/-) or homozygous (-/-) for the mouse ortholog of the human *DMK* gene, *dmk*, to implicate abnormal Na+ channel function in causing DMK deficiency in DM1. Saveliev *et al.*, (2003) have shown, using DMK transgenic mice, that triplet repeat expansions may result in gene silencing which in turn, may play a role in gene regulation. Other possible mechanisms associated with decreased DMK activity include impaired glucose utilisation, possibly related to abnormal protein kinase function, a potential relationship between an expanded DMK and an adjacent gene and an abnormal coronary reserve (Annane *et al.*, 1994; Annane *et al.*, 1996). It has also been proposed that the triplet repeat expansion within the *DMK* gene affects and interferes with the chromosome structure and with the expression of multiple neighbouring genes, resulting in the multisystemic clinical phenotype of DM (Jansen *et al.*, 1992; Mahadevan *et al.*, 1993; Perryman *et al.*, 1993).

Additionally, a number of families were found with clinical features characteristic of DM1, but without the CTG repeat in the *DMK* gene (Thornton *et al.*, 1994). Subsequently, a second locus for another DM-causing gene, DM2, was mapped to chromosome 3q21 (Ranum *et al.*, 1998). DM2 is caused by an expansion of a tetranucleotide (CCTG) repeat motif located in intron 1 of the zinc finger protein 9 gene (*ZNFP9*), which causes the same multisystemic features as DM1 (Liquori *et al.*, 2001). Parallels between these two types of diseases indicate that the microsatellite expansions in the unspliced DMK mRNA may be pathogenic and cause the disease (Liquori *et al.*, 2001). Ranum and Day (2004) have shown that both DM1 and DM2 are caused by a gain-of-function RNA, in which the CUG and CCUG repeats alter cellular function, including alternative splicing of various genes.

Experimental work using murine mutants indicate that mutant RNA is retained in myonuclei where it aggregates and interferes with the export of the wild type mRNA (Taneja *et al.*, 1995; Davis *et al.*, 1997). Another interesting finding was that a null murine mutant for the *dmk* gene (-/-) developed only cardiac conduction defects and did not show any other signs of DM, such as myotonia or cataract formation (Tappscott *et al.*, 2001; Burton and Davis, 2002). This study seems to imply that *dmk* is specifically involved in the cardiac conduction pathway and could possibly give indications about the mechanism involved in the development of PFHBI.

1.3.2.2 Limb girdle muscular dystrophy (LGMD)

Limb girdle muscular dystrophy constitutes a heterogeneous group of disorders characterised by a general limb-girdle weakness. A recently recognised autosomal dominant limb girdle muscular dystrophy 1B (LGMD1B:OMIM 159001), with a high degree of cardiac involvement, was mapped to chromosome 1q11-21 (Kass *et al.*, 1994). In affected individuals, symmetric weakness starts in the proximal lower-limb muscles before the age of 20 years, and in the third or fourth decade, the upper-limb muscles become affected. Van der Kooi *et al.*, (1997) described three new families with LGMD1B with cardiac involvement. Cardiac abnormalities found in 62.5% of the patients, included A-V conduction disturbances similar to the family described by Kass *et al.*, (1994). The disease-causing gene was found to be lamin A/C, *LMNA,* which encodes two proteins of the nuclear envelope. Muchir *et al.*, (2000) identified three specific mutations in *LMNA* in all the affected members of each LGMD1 family. Since cardiac conduction abnormalities in LGMD1B are linked to mutations in *LMNA*, and PFHBI is a disease which affects the CCS, one can propose that a *LMNA* paralog on chromosome 19, if present, would be a good gene to screen for PFHBI-causative mutations.

1.3.2.3 Familial bundle branch block diseases

Morquio first reported familial BB heart diseases in 1901. At a later stage Combrink *et al.* (1962), described another family, which included five brothers with A-V conduction disturbances. Thirty years ago the pathological basis for BBB was poorly defined and the diagnosis was based solely on an ECG profile (Davies, 1971). In addition, the wide branching of the BBs over the septum has made adequate histological examination

difficult (Davies, 1971). Graber *et al.*, (1986) concluded from a study of familial CCS and myocardial disease that extensive biochemical investigations would provide clues to the molecular aetiology of these diseases. BBB occurs when an interruption of conduction in one of the main BBs occurs and the presence of a prolonged QRS complex is observed on ECG. The normal width of the QRS complex is 0.1 second, but, in the case of BBB, it is greater than 0.1 second (Goldberger and Goldberger, 1990). CHB occurs when a total cessation of conduction of the impulse through the A-V node and the bundle of His occurs; that is, atrial and ventricular depolarisations are completely independent. Complete heart block (CHB) is potentially life-threatening and is treated by implantation of a ventricular pacemaker (Goldberger and Goldberger, 1990).

Globally, a number of familial BB conduction diseases have been described which share similarities to PFHBI, although identified under different names (see table 1.1) (Brink *et al.*, 1994), such as as progressive cardiac conduction disease (PCCD), isolated cardiac conduction disorder (ICCD) and PFHBII.

1.3.2.3.1 Progressive familial cardiac conduction disease (PCCD)

Progressive cardiac conduction disease (PCCD:OMIM 113900) segregates in an autosomal dominant manner in a French family who had originally been diagnosed with Lenègre-Lev disease (Lenègre, 1964; Lev, 1964). Long-term follow-ups of several affected members demonstrated that their conduction defect increased in severity with increasing age. Schott *et al.*, (1999) excluded linkage of PCCD to the PFHBI locus on chromosome 19, and other loci associated with inherited cardiac diseases with conduction

Table 1.1 Summary of familial bundle branch disease with similarities to PFHBI.

Table 1. Familial bundle branch disease. Summary of reports in the literature of disorders with similarities to PFHB.

Source		SB	RBBB	BFB	CHB	No. individuals: Pedigrees	Total/study	AFF
Moseuti	1954 (9)	–	+	–	–	1	11	4
Trivella *et al.*	1960 (10)	–	+	–	–	1	7	5
Combrink *et al.*	1962* (11)	–	+	–	+	1	6	4
Gazes *et al.*	1965 (12)	+	+	–	+	1	21	9
Simonsen *et al.*	1970 (13)	–	+	–	+	1	28	5
Steenkamp	1972* (14)	–	+	+	+	1	17	6
Myburgh *et al.*	1973* (15)	+	+	–	+	6	48	22
Schaal *et al.*	1973 (16)	–	+	+	+	1	45	31
Vallianos *et al.*	1974 (17)	–	+	+	+	1	14	8
Esscher *et al.*	1975 (18)	–	+	–	–	2	69	20
Brink *et al.*	1977* (1)	+	+	+	+	1	55	31
Stéphan	1979 (19)	–	+	+	+	3	505	128
v.d. Merwe	1988* (4)	+	+	+	+	many	181	84

*, South African families suffering from PFHBI; SB, sinus bradycardia; RBBB, right bundle branch block; BFB, bifascicular block; CHB, complete heart block; AFF, affected; +, ECG abnormality considered present; -, ECG abnormality not recorded (Brink *et al.*, 1994).

defects, and the PCCD locus was eventually mapped to chromosome 3p21. It was found that the disease-causative mutation occurred in the cardiac-specific Na^+ channel 5A gene (*SCN5A*), namely, a T to C transition in the highly conserved donor splice site of intron 22, resulting in the formation of an abnormal transcript with an in-frame skipping of exon 22 (Schott *et al.*, 1999). Thereafter, a second mutation was observed in a Dutch family, and in this case, the mutation was a 1bp deletion (G) at nucleotide 5280 of *SCN5A*, resulting in a frameshift causing a premature stop codon (Schott *et al.*, 1999). Interestingly, *SCN5A* has been implicated in three other cardiac conduction diseases, namely, Brugada syndrome, in which patients have marked abnormalities in ventricular

conduction (Vatta *et al.*, 2000), isolated cardiac conduction disease (ICCD)(section 1.3.2.3.2), a sustained, isolated conduction defect which slows down the cardiac rhythm (Tan *et al.*, 2001) and long QT syndrome-3 (LQT3) (Wang *et al.*, 1995).

1.3.2.3.2 Isolated cardiac conduction disease (ICCD)

In 2001, Tan *et al.*, found another mutation in *SCN5A,* which co-segregates with another conduction disorder called isolated cardiac conduction disease (ICCD). The ECG of a 3-year old patient with ICCD showed characteristics of slow conduction through the atria and ventricles, including a broad P wave, PR interval prolongation and a wide QRS complex. The described mutation in exon 12 of *SCN5A* results in a substitution of cysteine 514 for glycine (G514C) in the channel protein (Tan *et al.*, 2001). Recent work by Tan *et al.*, (2002) showed that a Ca^{2+} sensor in *SCN5A*-encoded channel modulates cardiac excitability and that a different mutation in the carboxy-terminal 'IQ' domain of the same protein causes ventricular arrhythmias. These findings support the notion that ion channels and or calcium homeostasis may play a role in the development of PFHBI.

1.3.2.3.3 Progressive familial heart block II (PFHBII)

Progressive familial heart block II (PFHBII:OMIM 140400) segregates in a South African family and is inherited in an autosomal dominant fashion (Brink and Torrington, 1977). It is characterised by a sinus bradycardia with a left posterior hemiblock, and a narrow QRS complex on ECG. Clinically, PFHBII presents with syncopal episodes and Stoke-Adams seizures when CHB supervenes and the conduction abnormalities are accompanied by dilated cardiomyopathy (Fernandez *et al.*, 2004). Recent linkage studies

have placed the PFHBII locus on the long arm of chromosome 1q32 (personal communication, Fernandez, 2004) although the disease-causing gene has not yet been identified.

1.3.2.3.4 Progressive familial heart block I (PFHBI)

Background

Progressive familial heart block I (OMIM 113900) was first described by Brink and Torrington in 1977 in several branches of a large South African pedigree. The disease is an autosomal, dominantly inherited, progressive BB conduction disorder that may progress to complete heart block (Brink and Torrington, 1977; Van der Merwe *et al.*, 1986; Van der Merwe *et al.*, 1988). In earlier studies, it had been estimated that there may be between 1000 and 9000 gene carriers among the descendants from a Portuguese ancestor (Brink and Torrington, 1977; Torrington *et al.*, 1986; Van der Merwe *et al.*, 1994). To date, three apparently unrelated pedigrees designated 1, 2 and 5 have been established (fig. 1.6). Pedigree 2 consists of nine generations, all descended from one ancestor who emigrated from Lisbon, Portugal in 1696 (Brink and Torrington, 1977). He married a woman of French origin in 1735 and they moved to the Eastern Cape near the Gouritz River. Pedigrees 1 and 5 are smaller pedigrees with the affected members displaying the same phenotype and disease-associated haplotype as individuals in pedigree 2. However, no family data linking pedigrees 1 and 5 to the same ancestor of pedigree 2 is available (Brink, 1997). Currently, a large branch of the family lives in the Free State Province, while others have moved to various places in South Africa (fig. 1.7).

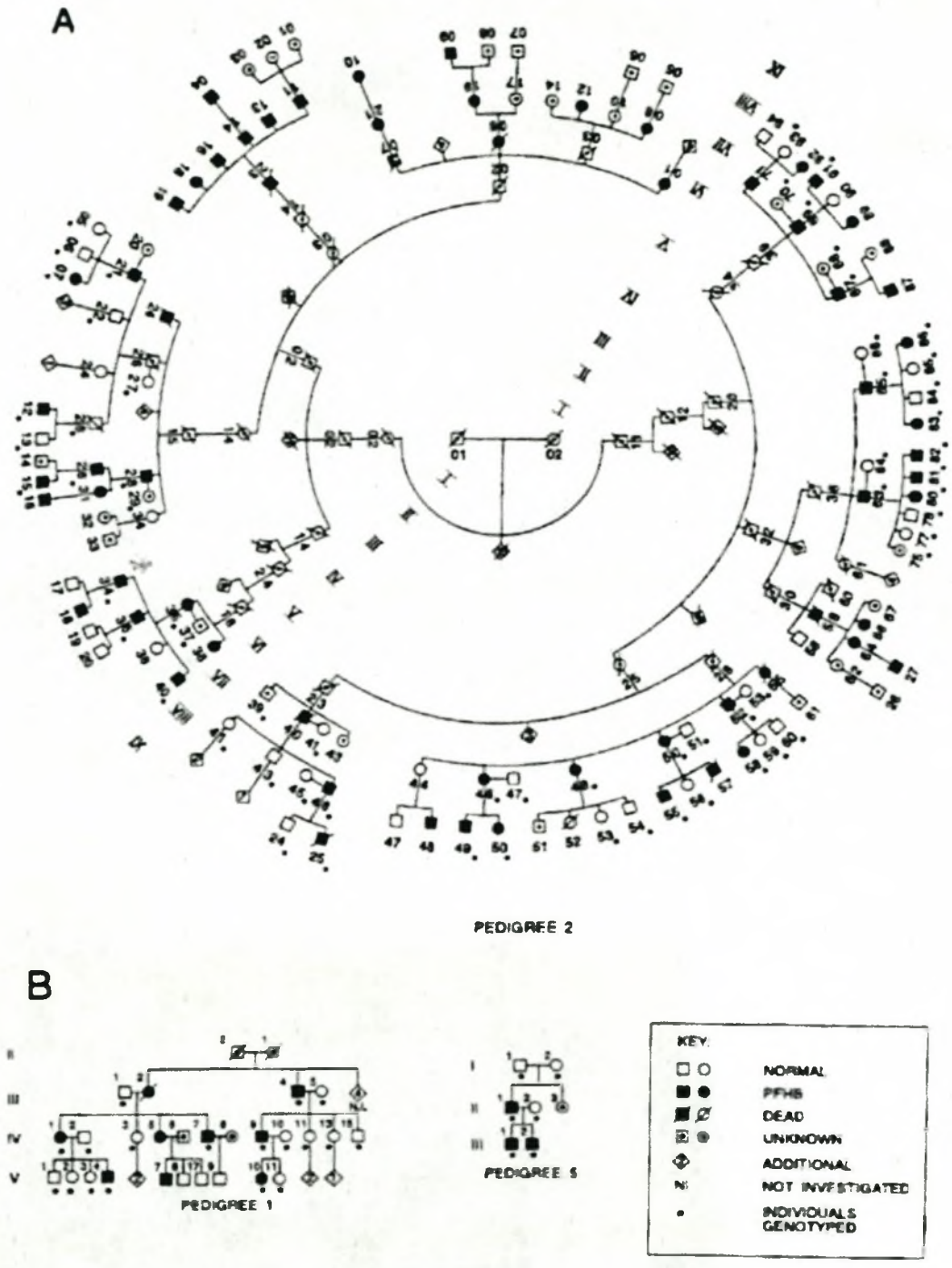


Figure 1.6 Family pedigrees 1, 2 and 5 segregating for PFHBI.
A, pedigree 2 is a nine generation pedigree traced back to a single male (I.01) and his wife (I.02). Individual I.01, was one of 10 siblings, all offspring of a Portuguese immigrant and his wife of French ancestry, who married in South Africa in 1735. B, pedigree 1 and 5, are two smaller pedigrees with affected members showing the same phenotype as individuals in pedigree 2 and were, therefore, regarded as segregating for PFHBI. No genealogical links have been established with pedigree 2. Affected individuals had either right bundle branch block, complicated right bundle branch block, or complete heart block and the same haplotype (Brink *et al.*, 1995).

Figure 1.7 The distribution and movement of the pedigree 2 PFHBI family in South Africa. →, indicating the movement of the PFHBI families from Cape Town to the Gouritz river, spreading to the Eastern Cape and then finally moving to the Free State and onto Gauteng (modified map from Encarta 2001).

PFHBI is defined on ECG by evidence of BB disease, i.e., RBBB, LAHB or LPHB or CHB with broad QRS-complexes (fig. 1.8) (Brink and Torrington, 1977).

Age of onset

There are four high-risk periods reported for the onset of PFHBI, namely at birth, soon after birth, the early 20s, and again towards middle age (Brink and Torrington, 1977). It was also found that in familial cases an early normal ECG does not exclude the occurrence of an underlying conduction disorder in later life (Brink and Torrington, 1977).

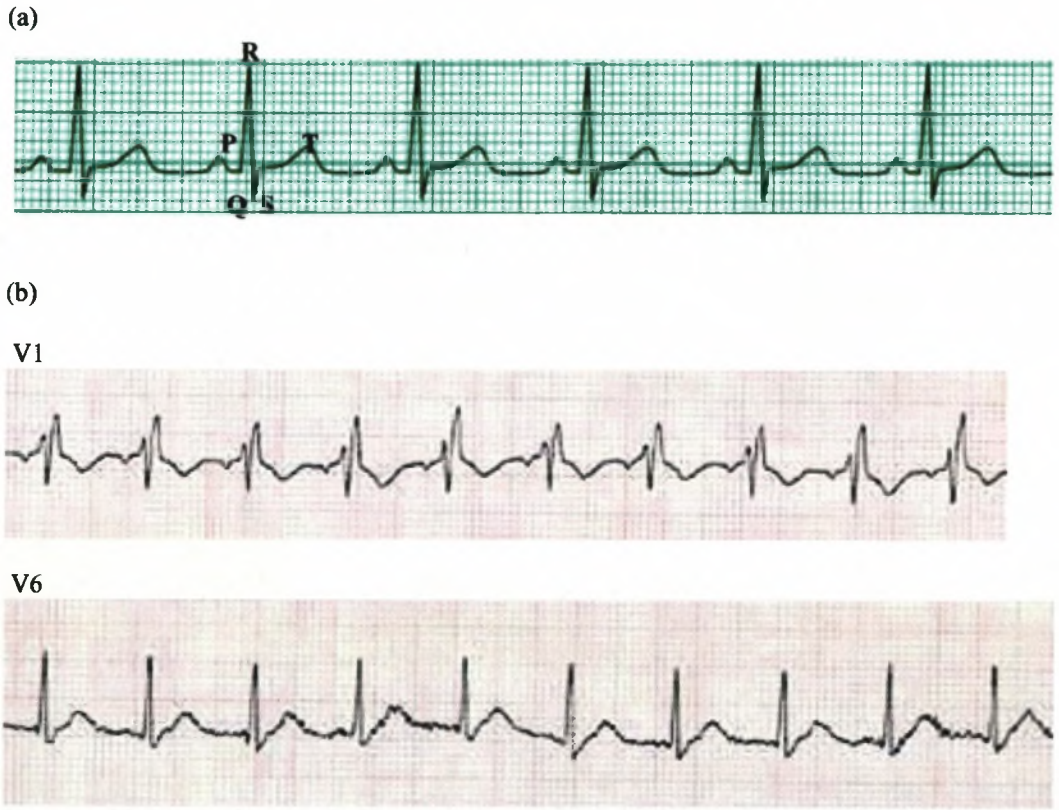


Figure 1.8 ECG tracings showing a typical pattern of right bundle branch block and a normal ECG. (a) A typical ECG showing the deflections resulting from the electrical changes in the components of the conduction system resulting in the P, Q, R, S and T waves. In a normal individual, the QRS complex is 120 ms wide. (b) An ECG obtained from an individual with right bundle branch block. The QRS complex is more than 120 ms wide with a delayed terminal wave attributable to late right ventricular depolarisation as shown by a terminal slurred R wave in standard V1 and a slurred S wave in V6 (http://www.lond.ambulance.freeuk.com/ecg/ECG2htm).

These findings suggest that the PFHBI-causative gene might be involved in the developmental pathways of the CCS, since the manifestation of the disease occurs at specific developmental stages.

Pathology of PFHBI

By way of comparison, pathological studies of the CCS, in DM1-affected individuals showed evidence of non-specific conduction fibre loss throughout the system, with replacement by fat rather than fibrosis (Nguyen *et al.*, 1988). The only autopsy from a PFHBI-affected individual was performed on a 2 year-old child, and showed fibrosis of the conduction tissue, namely, the main and the proximal part of the BBs, in the presence of dilated cardiomyopathy. However, the post mortem also revealed that the patient died of a cerebral embolus (Van der Merwe *et al.*, 1991), which was considered atypical of PFHBI.

Replacement fibrosis is a common feature in general conduction disease (Davies *et al.*, 1983). The process involves a dying-off of the BB tissue, with substitution of fibrotic tissue. Post-mortem studies by Davies in 1971 showed that 46% of patients suffering from CHB had fibrosis of the bundle of His and BBs in the absence of any other myocardial disease. The fibrosis of the BBs might be caused by an underlying collagen defect termed "exaggerated ageing effects", as suggested by Lev in 1964, or "primary degenerative Lenègre disease" as suggested by Lenègre (1964).

Currently, the only treatment for PFHBI is the timely implantation of a pacemaker, which enables the affected individual to lead a normal and productive life (Torrington *et al.*, 1986).

Anticipation

The phenomenon of anticipation was considered to be present within the PFHBI pedigrees, as it had been noted that successive generations required earlier implantation of pacemakers (Van der Merwe *et al.*, 1988). The mean age of pacemaker implantation in PFHBI-affected individuals decreased with successive generations, namely, implantation was at 9,8 years in generation 8, 30,8 years in generation 7 and at 60,5 years in generation 6; suggesting that members of the younger generation are more severely affected than members of the older generation. Unusual triplet and tetranucleotide repeat expansion have been associated with a number of diseases which show anticipation (Sheffield *et al.*, 1995; Rosenberg, 1996; Liquori *et al.*, 2001 (section 1.4.1.3). As part of this study, the search for expanded triplet repeats was pursued.

Linkage studies to identify the PFHBI locus

Brink *et al.*, (1994) performed a linkage study using pedigrees 1 and 2 originally reported by Brink and Torrington (1977). Of 80 marker loci tested, 68 were excluded from linkage with the PFHBI locus. Additionally, large portions of chromosome 1 (72%), chromosome 4 (70%) and chromosome 19 (55%) were excluded from linkage using multipoint mapping (Brink *et al.*, 1994). Subsequent linkage studies mapped the PFHBI gene to a region of approximately 10cM on chromosome 19q13.2-q13.3 (Brink *et al.*, 1995).

A French group mapped a second conduction defect named ICCD by linkage analysis to the same region of chromosome 19q13.3 as PFHBI (De Meeus *et al.*, 1995). Although ICCD has the same name as a disease identified by Schott *et al.*, (1999) and Tan *et al.*, (2001), it has different clinical features and is caused by different chromosome mutations in the *SCN5A* gene on a different chromosome too. To avoid confusion, it will be referred to as CCD in this thesis. Through collaborative studies conducted by the PFHBI group with the French group, it was confirmed that common clinical features are shared by both PFHBI and CCD. On this basis, it was decided that these two conduction disorders are likely to be the same disease (personal communication, Brink and Bouvagnet, 1998).

The PFHBI locus centered on the kallikrein (*KLK1*) locus, within a 10cM region and was flanked by the apolipoprotein C-II *(APOC2)* and related RAS viral (r-ras) oncogene (*RRAS)* genes (Brink *et al.*, 1995), whereas the CCD gene was localised to a larger 13cM interval, between the flanking markers *D19S606* and *D19S571* (De Meeus *et al.*, 1995) (fig. 1.9).

An integrated map combining the different markers used by the two groups showed that the two intervals overlapped. Using genetic fine mapping, the PFHBI locus was reduced to 7cM between markers *D19S412* and *D19S866* (fig. 1.9) (personal communication, De Jager, 1998). As part of the present study, fine genetic mapping was continued to further reduce the area in order to localise the disease-causing gene.

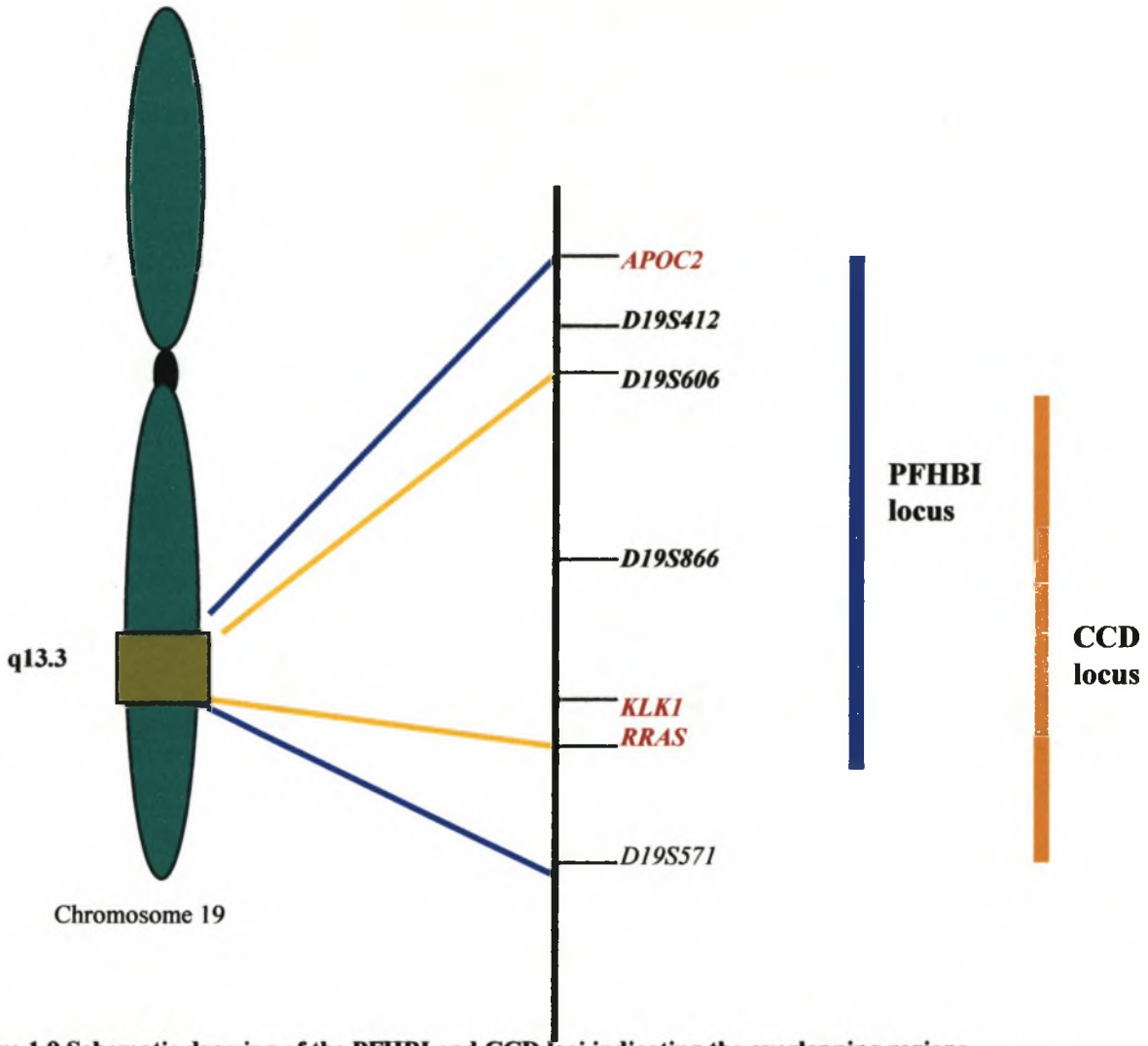


Figure 1.9 Schematic drawing of the PFHBI and CCD loci indicating the overlapping regions. The solid blue lines indicate the PFHBI locus. The solid orange lines indicate the ICCD locus. *D19S412, D19S606, D19S866* and *D19S571* are genetic markers. *APOC2*, Apoliprotein C-II; *RRAS*, related RAS viral (r-ras) oncogene and *KLK1*, kallikrein 1 are genes flanking the PFHBI locus (Brink *et al.*, 1995; de Meeus *et al.*, 1995).

1.4 POSITIONAL CANDIDATE STRATEGIES

The Human Genome Project (HGP) has generated a number of resources, such as genetic, physical and transcript maps, DNA clones, sequences, expression and phenotypic data, which have sped the identification of novel disease-causing genes. As the HGP progressed, search strategies were modified so that available data from the publicly

accessible databases and bioinformatic tools could aid the search for the PFHBI gene. Some of the strategies used to find the PFHBI gene are described in sections 1.4.1, 1.4.2 and 1.4.3 of this study, and include linkage analysis, genetic fine mapping, somatic cell hybrid mapping, bacterial artificial chromosome (BAC) and cosmid mapping, establishing of clone contigs, and retrieving and clustering of expressed sequence tags (ESTs). An EST is a partial sequence of a fragment of a cDNA clone, generated from 5'- and 3'-ends of the clone, that is between 150 to 200bp in length and represents an uncorrected single read sequence (Adams *et al.*, 1991) (see section 1.4.3.1 for a more detailed discussion of ESTs).

1.4.1 Genetic map

The earliest classical human genetic maps consisted of protein polymorphisms, notably blood groups and serum protein markers, which are both rare and not very informative (Gyapay *et al.*, 1994). In 1987, the first complete genetic map of the human genome was published using restriction fragment length polymorphisms (RFLPs) (Donis-Keller *et al.*, 1987). The discovery of microsatellites allowed great advances in the development of a human genetic map and aided its resolution (Weber and May, 1989). Subsequently, Weber *et al.*, (1993) generated a higher resolution linkage map of short tandem repeat (STR) markers.

Since the 1990's, there have been rapid advances in developing genetic maps for humans, greatly advancing the ability of researchers to localise and identify genes for inherited disorders. Through the efforts of three large groups generating microsatellite markers,

and the efforts of 110 collaborators of the Centre d'Études du Polymorphisme Humain (CEPH), comprehensive linkage maps of all the human chromosomes are now available (Cooperative Human Linkage Centre, CHLC, 1994). The Généthon group in France worked on maps based on markers that contained CA repeat motifs, which are the most common STR in the human genome (Weissenbach *et al.*, 1992; Gyapay *et al.*, 1994). The CHLC and the Utah marker development group (based in the state of Utah, USA) developed maps based on di-, tri-, and tetranucleotide repeats (CHLC, 1994). The maps were produced through the collaborative efforts of the different groups, which generated markers at an average density of 0.7cM (CHLC, 1994). Searches for disease-causing genes require the establishment of fine genetic maps and saturation fine mapping using of polymorphic markers, to detect recombination events (Keating, 1992), as in the case of the search for the PFHBI gene.

During the course of the PFHBI gene search, the available genetic markers of the current STR map of chromosome 19q13.3 had been utilised and the development of novel STRs was required for further fine mapping studies of the PFHBI locus. This involved the search for di-, tri- and tetranucleotide sequences in DNA clones spanning the PFHBI locus (Christoffels, 1997; Makubalo, 2000).

1.4.1.1 Dinucleotide markers

$(CA)_n$ repeat sequences are abundantly dispersed repetitive elements present in genomes of many eukaryotes, including humans (Hamada *et al.*, 1984) and are the most frequently found repeat sequences in the human genome (Beckman *et al.*, 1992). In 1989, Tautz

demonstrated hypervariability in the length of the $(CA)_n$ repeat sequences of two loci in a family, in addition, the repeats showed a Mendelian pattern of inheritance. The polymorphic and Mendelian characteristics of these motifs were developed as a source of genetic markers (Gyapay *et al.*, 1994).

There are three categories of $(CA)_n$ repeat sequences, namely, perfect repeats without interruptions in the runs of CA, imperfect repeats with one or more interruptions in the run of repeats and a compound of repeat sequences with adjacent tandem simple repeats of a different sequence (Weber, 1990). Weber (1990) demonstrated that loci with 10 or fewer $(CA)_n$ repeats had little or no allelic variation, whereas loci with 11 to 15 repeats were more polymorphic and those with 16 or more repeats were highly polymorphic and therefore highly informative markers. The function of $(CA)_n$ repeats is unknown, but it has been proposed that they serve as hot spots for recombination (Slightom *et al.*, 1980) or participate in gene regulation (Hamada *et al.*, 1984).

1.4.1.2 Tri- and tetranucleotide markers

Human trimeric and tetrameric short tandem repeats (STRs) are highly polymorphic in humans, where they are distributed randomly within the intronic and sometimes in the exonic regions of the genome on an average of every 300kb (Edwards *et al.*, 1992) and are generally inherited stably (Edwards *et al.*, 1991). It was found that the informativeness of a STR was dependent on the repeat type and length e.g., STRs with eight or more triplet repeat motifs (Gastier *et al.*, 1995) and tetranucleotide repeats of six or more are informative (Sheffield *et al.*, 1995). The tetranucleotide STRs occurring most

abundantly are AAAG (38%), AGAT (29%), AAGG (10%) and AAAT (5%) (Utah Marker Development Group, 1995).

However, a different class of trinucleotide and tetranucleotide repeats has been found which are distributed in both intronic and exonic regions, which may sometimes be pathogenic (section 1.3.2.1 and 1.4.1.3) and which are not stably inherited (Chakraborty *et al.*, 1997).

1.4.1.3 Pathogenic microsatellites

At least 13 inherited diseases are caused by mutations in triplet repeat sequences, characterised by the abnormal expansion of the mutated allele (Warren, 1996). For example, in the case of DM1, the normal allele is 5-30 CTG repeats in length and the mutated allele is expanded to 50-2000 in the affected individual (Groh *et al.*, 2002).

A number of different mechanisms have been proposed to explain how the expansion of trinucleotide repeats may cause disease. Suggested mechanisms are interference with either transcription, a gain of function if exonic or a protein conformation change (Jennings, 1995). In the case of Friedreich's ataxia (FRDA), the abnormally expanded GAA repeat adopts a triple helical conformation at the DNA level *in vitro*, which may be responsible for the suppression of gene expression (Ohshima *et al.*, 1998). Interestingly, DM1 and DM2 are caused by a non-translated trinucleotide (CTG) (Brook *et al.*, 1992; Mahadevan *et al.*, 1992) and tetranucleotide (CCTG) expansion (Liquori *et al.*, 2001), respectively. The mechanism by which these amplified repeat expansions lead to DM is discussed in section 1.3.2.1.

Since anticipation was reported in PFHBI-affected families (Van der Merwe *et al.*, 1986), the exonic sequences of genes within the PFHBI locus containing triplet repeats were examined. In previous studies, possible pathological expansion of the GAT and the GAG triplet repeat in exon 1 of histidine-rich calcium binding protein (*HRC)* (Christoffels, 1997) and the expansion of a CAG triplet repeat in exon 13 of nucleobindin 1 (*NUCB1)* (*F. February, 2002) were investigated, but no expansion of the repeat motifs was found in these genes in PFHBI-affected individuals.

1.4.2 Chromosome 19

Human chromosome 19 is estimated to be 55.8 Megabases (Mb) in length (Grimwood *et al.*, 2004) and constitutes slightly more than 2% of the human genome (Mayall *et al.*, 1984). Previously, the International Human Genome Sequencing Consortium (IHGSC), (2001) determined the chromosome size to be 72Mb, which included all gaps and duplications. Chromosome 19 is known to have a relatively high GC content (Larsden *et al.*, 1992), estimated at an average of 48% (fig. 1.10) (Grimwood *et al.*, 2004) compared to 41% for the whole human genome (IHGSC, 2001). Although chromosome 19 is relatively small compared to the other chromosomes, it has a very high density of genes, namely, 26 protein-encoding gene loci per Mb (Grimwood *et al.*, 2004), making it the most gene-rich chromosome of all the chromosomes (Venter *et al.*, 2001). In total, 1,461 protein encoding genes and 321 pseudogenes have been placed on chromosome 19.

* Please note that when a reference is made to F. February in the text, it refers to an individual.

Chromosome 19 is also unusual in its density of repeat sequences, because nearly 55% of this chromosome consists of repetitive elements, whereas chromosomes 6, 7, 14, 20, 21 and 22 all have repeat contents ranging from 40% to 46% (IHGSC, 2001). *Alu* repeats make up 25.8% of the chromosome compared to 13.8%, 13.3%, 9.5% and 16.8% on chromosomes 7, 14, 21 and 22, respectively (Grimwood *et al.*, 2004). It has been estimated that 568 of the estimated 1,461 genes (39%) show evidence of alternative splicing. In addition, 321 pseudogenes have been identified, of which 177 (55%) are classified as "processed" pseudogenes, that is, products of viral retrotransposition events involving spliced mRNA. It has also been found that chromosome 19 is notable for the prevalence of duplication, namely, tandemly clustered gene families and large segmental duplication with evidence of genomic duplication. Sequence homology of 7.35% of chromosome 19 sequences are shared with more than one location within the genome. Finally, comparative analyses revealed large blocks of genes with orthology to rodent genes and segments of coding and non-coding sequence conserved in the fish species *Takifugu* (fig. 1.10) (Grimwood *et al.*, 2004).

As part of the HGP, the sequencing of chromosome 19 was assigned to the sequencing group based at the Human Genome Centre at the Lawrence Livermore National Laboratories (LLNL), Livermore, California, which resulted in the generation of the first published physical map of chromosome 19 (Ashworth *et al.*, 1995) (see section 1.4.2.2). Subsequently, the final map containing the full chromosome 19 sequence was published in 2004 (Grimwood *et al.*, 2004).

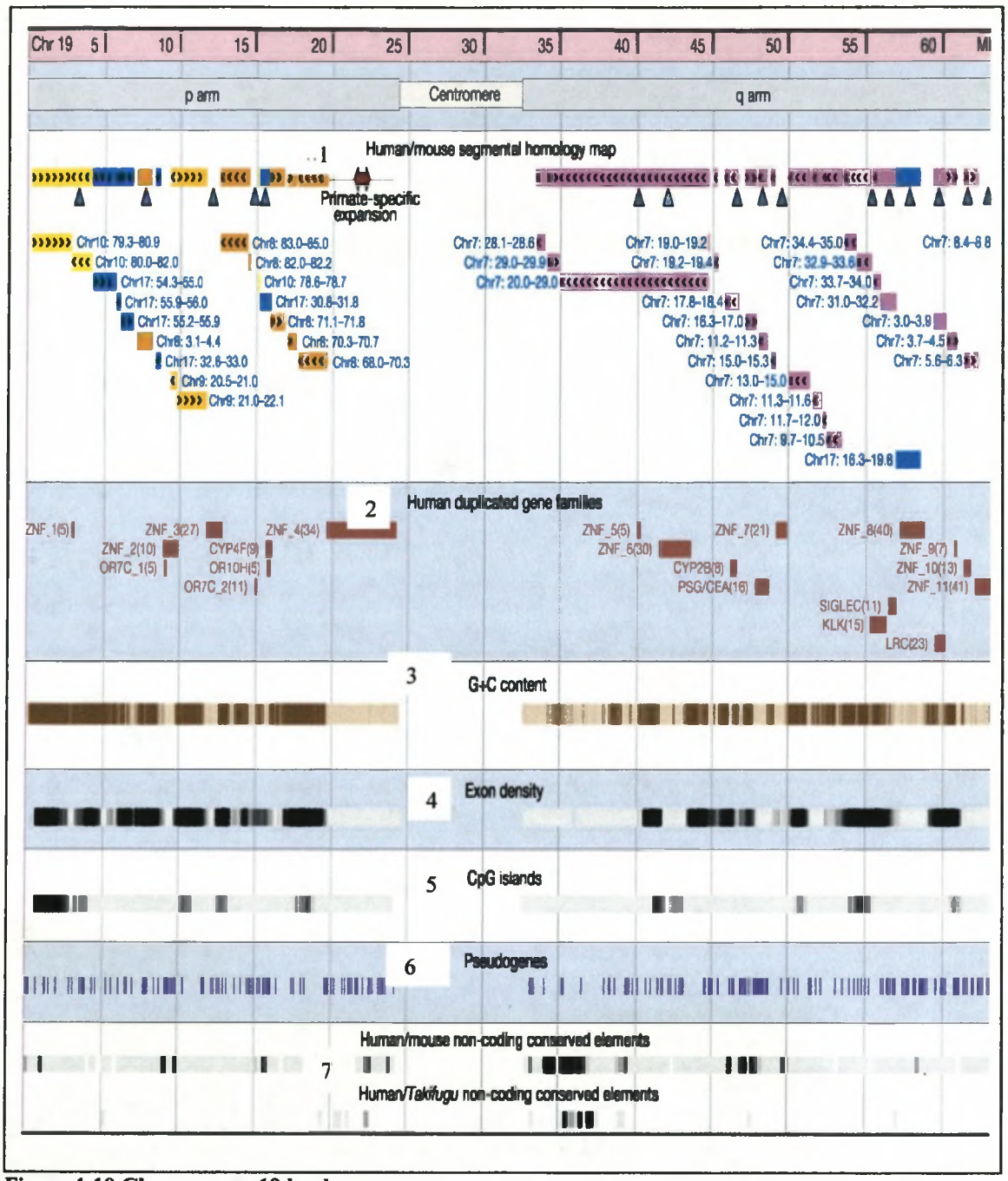


Figure 1.10 Chromosome 19 landscape.
The following sections appear from top to bottom: (1) The human and mouse blocks of homology larger than 100kb are shown. The grey triangles indicate regions where significant lineage-specific and orthology can be identified. Extensive intrachromosomal and interchromosomal rearrangement between human and mouse can be observed. (2) The duplicated gene families cover more than 25% of the chromosome. (3) The G+C content profile shows that the zinc finger and olfactory receptor gene families have a lower G+C content than the rest of the chromosome which has a very high G+C content. (4) The exon density correlates with the G+C content fluctuations. (5) The location of the CpG islands on chromosome 19. (6) The psuedogenes identified during the annotation of chromosome 19. (7) Human/mouse and human/*Takifugu* non-coding DNA element density. This density is unevenly distributed over the chromosome with a 5Mb region in the proximal region of the q arm containing the highest density of human/mouse non-coding and the majority of the non-coding of human/*Takifugu* elements (Grimwood *et al.*, 2004).

1.4.2.1 Somatic cell hybrids

Somatic cell hybrids are produced by the fusion of human and rodent cell lines under specific experimental conditions. Several somatic cell hybrids for the physical mapping of chromosome 19 have been described (Brook *et al.*, 1986; Hulsebos *et al.*, 1986). Schonk *et al.*, (1989) constructed and characterised a series of somatic cell hybrids of chromosome 19q, which allowed a finer physical subdivision of the region (fig. 1.11). Bachinski *et al.*, (1993) constructed a panel of 22 somatic cell hybrids, which divided the chromosome 19q arm into 22 ordered subregions. The panel was characterised with respect to 41 genetic markers and it aided in the establishment of the order of a number of genes that had already been placed in this region of chromosome 19q.

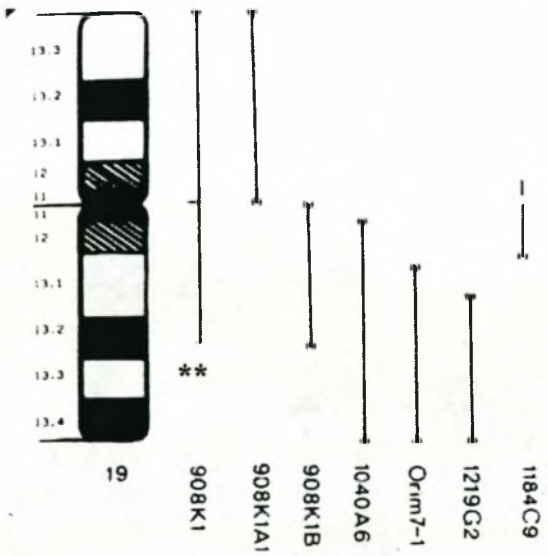


Figure 1.11 Diagram of chromosome 19 segments contained in somatic cell hybrids.
908K1, 908K1A1, 908K1B, 104016, Orim7-1, 1219G2 and 1184C9, somatic cell hybrids; **, position of the PFHBI locus (Schonk *et al.*, 1989).

1.4.2.2 Metric FISH framework map of chromosome 19

Ashworth *et al.*, (1995) published the first metric physical map of human chromosome 19 (fig. 1.12) which was composed of overlapping cosmids contigs, generated by automated fingerprinting, spanning 50Mb. The distances between selected cosmid clones were estimated using FISH, which provided order and distance between contigs. FISH maps were constructed of the p-arm (Brandriff *et al.*, 1994) and the q-arm (Gordon *et al.*, 1995) with an average inter-marker distance of 230kb across the non-centromeric sections of the chromosome. An additional 57 cosmids were ordered relative to the known reference clones. The published map of chromosome 19 consisted of 51 'islands' containing information from cosmid, YAC, BAC, and P1 artificial chromosome (PAC) clones, whose size and order were known. Over 450 genes, genetic markers, sequence tagged sites (STSs), anonymous cDNAs and other markers were placed on the chromosome 19 map. Figure 1.13 shows the DNA clones, genes, genetic markers, STSs and cDNAs at the PFHBI locus in February 2003.

In addition, a *EcoRI* restriction map was generated for 41Mb (~83%) of the chromosome which enabled the selection of minimal overlapping clones for DNA sequencing. This map, however, contained a number of gaps between contigs, and consequently, mainly BACs and, in some instances, PACs were used to effect gap closure. The cosmid contigs were extended and merged by hybridisation screening of cosmid and BAC libraries using a number of different probes, such as *Alu*-polymerase chain reaction (PCR) products, sequence tagged site (STS) PCR products, oligonucleotides, and cDNAs (Ashworth *et al.*, 1995). This resulted in an in depth coverage of overlapping clone contigs, which

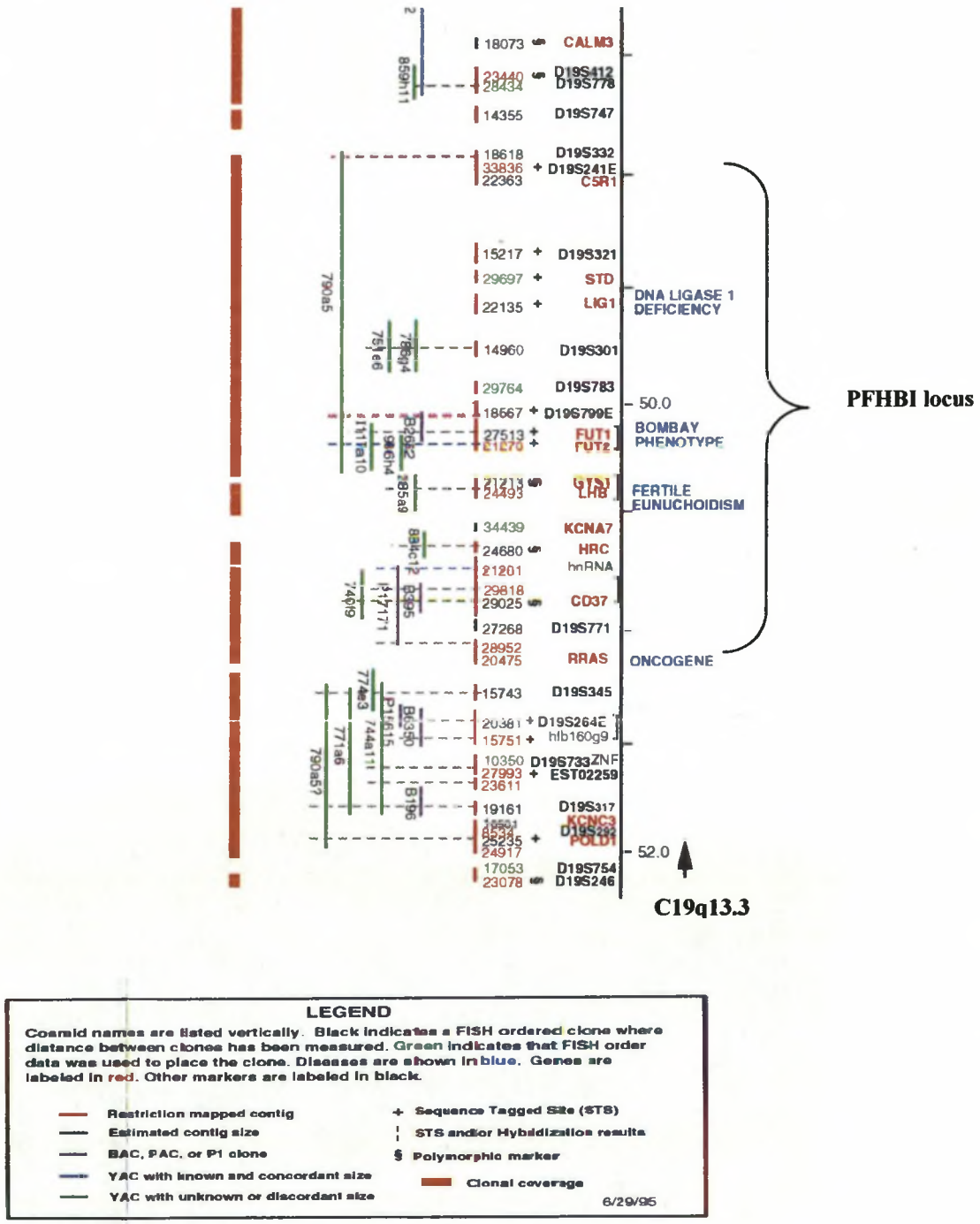


Figure 1.12 Contig map of the C19q13.3 region (Ashworth *et al.*, 1995).

facilitated the selection of a minimal number of clones for sequencing, reducing the cost and time of the project. In 1997, the Joint Genome Institute (JGI) was established and continued the sequencing of chromosome 19, in addition to chromosome 5 and chromosome16 (http://www.jgi.doe.gov). In 1999, a new Production Genomics Facility in Walnut Creek, California was established. The month of February 2001 heralded the publishing of a number of papers describing the initial sequencing and analysis of the human genome, including chromosome 19, as discussed in section 1.4.2 (IHGSC, 2001; Venter *et al.*, 2001). In 2002, the Stanford Human Genome Center based at Stanford University (http://www.shgc.stanford,edu/) continued the sequencing of chromosome 19 and generated 300Mb of finished human sequence. By April 2004, 99.9% chromosome of 19 was sequenced and the analysis of the most recent map was published (Grimwood *et al.,* 2004). The entire sequence for chromosome 19q13.3 is available in Genbank under the accession number NT_011109, which was used in this study.

1.4.3 Transcript map

The metric map generated by Ashworth *et al.*, (1995) was also used to refine placements of ESTs that had previously been assigned to somatic cell or radiation hybrid panels. At that stage of the project, very few ESTs had been placed on the map (fig. 1.12). The next transcript map of the region, spanning 10Mb on chromosome 19q13.3 to 19q.13.4, between markers *D19S408* and *D19S866* (Hamshere *et al.*, 1999), was derived using exon trapping to identify coding regions (Duyk *et al.*, 1990) and from genomic data deposited in the databases. Exon trapping was performed for this region of chromosome 19q on pools of three cosmids from contigs that had been anchored to the physical map

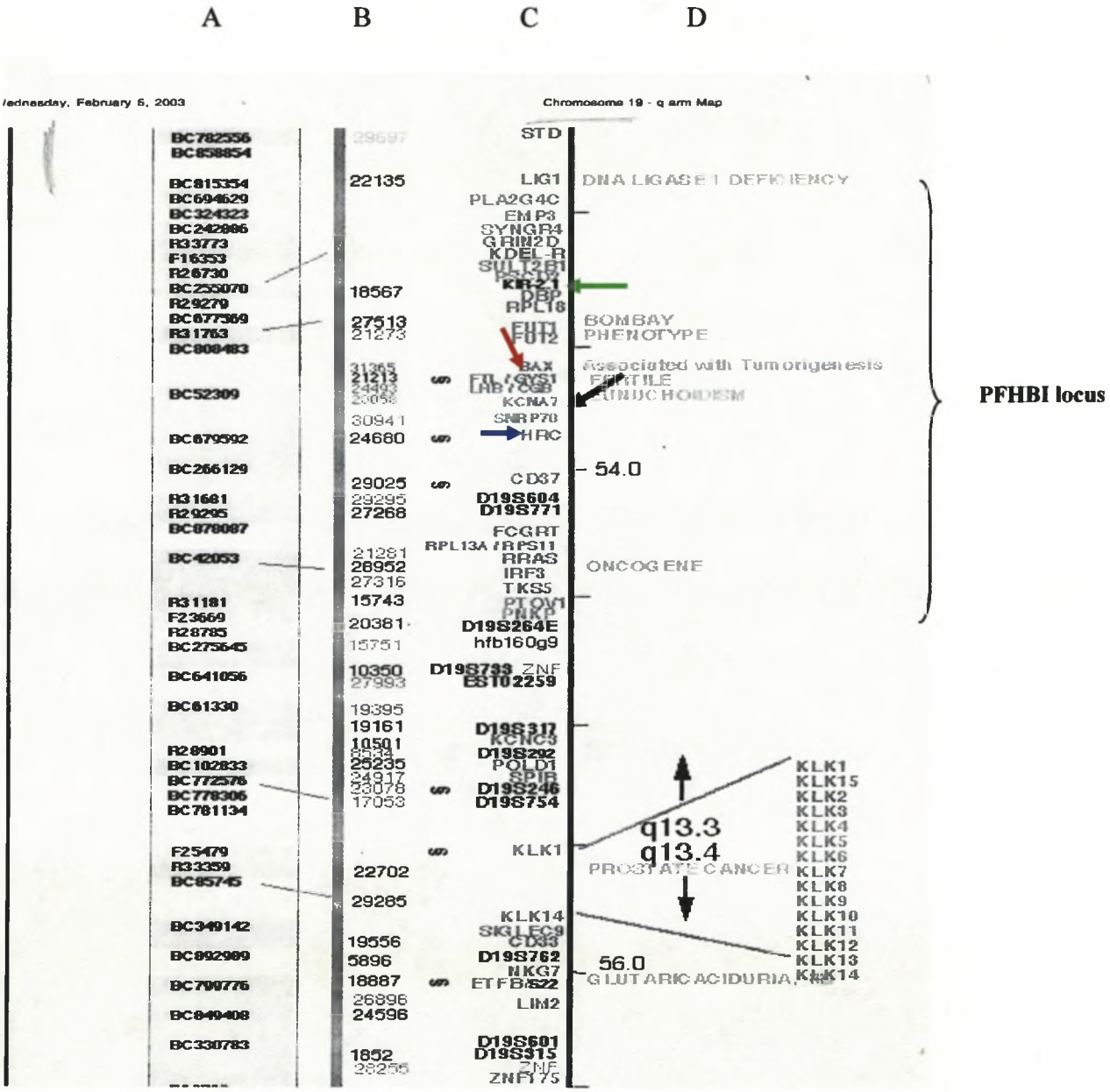


Figure 1.13 LLNL map showing the PFHBI locus (February 5, 2003)
Columns: A, BACs and cosmids spanning chromosome 19q; B, cosmids spanning the region, C, genetic markers, ESTs and genes mapping to the region, D, disease genes mapping to the region. **Red** arrow indicates the position of *BAX* and *GSY1*. **Black** arrow indicates the position of *KCNA7*. **Blue** arrow indicates the position of *HRC*. **Green** arrow indicates the position of *KIR2.4* (http://greengenes.llnl.gov/genome/).

from LLNL (http://www-bio.llnl.gov/bbrp/bbrp.homepage.html), which aided the identification of 96 unique exonic sequences. Of these, 55 matched EST sequences, which had been deposited in the databases at the United Kingdom (UK) Human Genome Mapping Project (HGMP) and NCBI. The remaining 41 exonic sequences, which did not show a significant match to sequences in the databases, were deposited at the EMBO alerting page (http://www.bork.embl-heidelberg.de/Alerting/).

1.4.3.1 Identification and characterisation of ESTs

With the advent of high-throughput technology and the need to hasten the search for disease-causing genes, large-scale partial sequencing cDNA projects, which generated a vast number of ESTs of all human chromosomes including chromosome 19, were undertaken. Hillier *et al.*, (1996) from the Genome Sequencing Center, Washington, USA generated 280, 000 human ESTs obtained from 194,031 human cDNA clones and 17 different tissues, representing three developmental states. The Integrated Molecular Analysis of Gene Expression (IMAGE) and the Institute of Genomic Research (TIGR) have also generated a large number of ESTs that have been deposited into the publicly available database, dbEST (Boguski and Schuler, 1995).

Cardiac ESTs

Clones from cardiac cDNA libraries at different developmental stages, namely, embryonic, foetal and adult, were partially sequenced to give an insight into gene involvement in cardiovascular disease. Liew *et al.*, (1994) generated 3,500 human cardiac ESTs and by performing a sequence similarity search against the Genbank/EMBL

databases, generated the following results: of the ESTs, 44.1% matched known genes, 47.4% had no match and the remaining 8.5% matched previously deposited human ESTs. Hwang *et al.*, (1997) from the Centre of Cardiovascular Research, Ontario, Canada, generated 43,285 ESTs from human heart cDNA libraries. This group retrieved an additional 41,619 ESTs from public databases. This total of 84,904 ESTs were obtained from 13 independent cardiovascular-based cDNA libraries, 55% of these ESTs matched known genes in Genbank/EMBL/DDBJ, 33% matched other ESTs, and 12% did not match any known sequence. ESTs matching known genes were classified according to function, transcription patterns of various tissue or developmental stage. In addition, over 57,000 ESTs from nine cDNA libraries constructed from human heart tissue from various anatomical, developmental and pathological stages were generated. In a review by Dempsey *et al.*, (2000), it was reported that the total number of ESTs deposited into Genbank from cardiovascular tissue represents less than 4%, of the total number from all tissues. This finding emphasises the need to increase the production of cardiovascular-based ESTs, which will improve the usefulness of this resource and allow better exploitation by cardiovascular researchers.

In addition, a wealth of information on the basic pathways of heart development can also be obtained by comparison of EST profiles from model organisms such as *D. melanogaster*, *X. laevis* and *D. rerio*. However, EST resources currently available for these organisms are severely limited, which greatly reduces the effectiveness of the approach (only 63,659 cardiac ESTs from cDNA libraries of *D. rerio* libraries, 22,788 ESTs from *M. musculus*-heart and 5,127 ESTs from *R. norvegicus*-heart) (reviewed by

Dempsey *et al.*, 2000). Subsequently, a vast number of additional ESTs from different organisms have been deposited into Genbank (Benson *et al.*, 2003).

1.4.3.2 Bioinformatic tools for EST analysis

Usually an EST is only 150-200bp long, which is too short for a significant protein homology search. In order to work with a longer sequence length EST, clustering programmes such as STACK at SANBI (Miller *et al.*, 1999; Christoffels *et al.*, 2001), or UNIGENE at NCBI (http://www.ncbi.nlm.nih.gov/unigene), are used. Once the ESTs are submitted to the respective programme, a longer consensus sequence, or cluster, is generated, which when BLAST-searched against protein databases is likely to produce a more meaningful protein hit than a shorter sequence.

Although both UNIGENE and STACK are EST clustering programmes, they have different EST data sources and use different mathematical processing methods, i.e., algorithms, to group sequences. Thus, UNIGENE utilises ESTs and only properly annotated mRNA sequences derived from the dbEST and Genbank databases (Boguski *et al.*, 1993; Benson *et al.*, 1996). The EST sequences are compared with the respective mRNA and all the sequences that share sequence homology with the mRNA, either at the 3'- or 5'-end, are grouped together. If the ESTs do not overlap with each other, they need to share homology to the template (mRNA) used and are placed into a single group, which represents a gene. In addition, sequences are also grouped if sequencing reads are performed from the same clone according to the template used. In some cases, no consensus sequence is found for a cluster (Bouck *et al.*, 1999). Regarding the STACK

dataset, EST data obtained from dbEST sequence data is classified according to the tissue from which it was derived. The clustering algorithm used identifies ESTs that are highly similar (>95% identical over 150 bases) and these sequences are aligned to provide a consensus sequence. At this point, the ESTs are grouped together if they are derived from the same cDNA clone. The way in which the two clustering systems handle problems, such as chimeric clones and alternatively spliced transcripts, are different, as will be discussed in the following paragraph.

Chimeric clones contain sequences from two different cDNAs that can cause serious problems when the clustering analysis is performed. UNIGENE handles the problem by searching for a single clone that joins two otherwise unconnected groups. Since STACK uses alignment in its system, chimeras are more easily identified during the alignment process.

The second problem is the identification of alternatively spliced transcripts, because these sequences contain identical regions, as well as disparate ones. As different isoforms of a transcript are generally expressed in different tissues, or at different stages of development (Bouck *et al.,* 1999), STACK separates transcripts by their origins; therefore, alternative transcripts should be significantly limited within a group. Whereas, with UNIGENE, fairly relaxed criteria are used for clustering; therefore groups containing transcripts with a small number of exons are common and, for this reason, the programme is not able to distinguish between chimeric clones and alternatively spliced transcripts (Bouck *et al.*, 1999).

EST_GENOME is a programme used for the alignment of spliced cDNA to unspliced genomic DNA (Mott, 1997). Other standard alignment tools are not able to correctly align a spliced product to genomic DNA, if large introns occur in the genomic sequence and conserved sequences found at intron/exon boundaries are ignored.

1.4.4 Morbid map of chromosome 19q13.3

The development of a more detailed map of chromosome 19q13.3 was aided because a number of different disease loci were mapped in or near the PFHBI target, namely, limb-girdle muscular dystrophy (LGMD2), Camurati-Engelmann disease (CED), a nonsyndromic deafness disorder (DFNA4) and DM1 (fig. 1.14).

LGMD2 is an autosomal recessively inherited disorder and belongs to a genetically heterogeneous group of diseases characterised by progressive degeneration of pelvic and shoulder muscles. The LGMD2 locus maps to chromosome 19q13.3 to a 9cM interval between genetic markers, *D19S412* and *D19S879,* within the PFHBI locus (fig. 1.14) (Driss *et al.*, 2000). Genomic screening identified an area of interest on chromosome 19, as evidence of linkage was observed with the microsatellite marker *D19S606*. In addition, genetic distances between microsatellite markers were given, namely, *D19S412*-6.3cM-*D19S606*-2.7cM-*D19S879*-2.3cM, which aided in mapping the PFHBI locus.

CED is an autosomal dominantly inherited disorder, which is characterised by the clinical features of a waddling gait, muscle weakness and wasting, and generalised fatigue. The CED locus maps to a 15.1cM region on chromosome 19q13.1-q13.3 and is

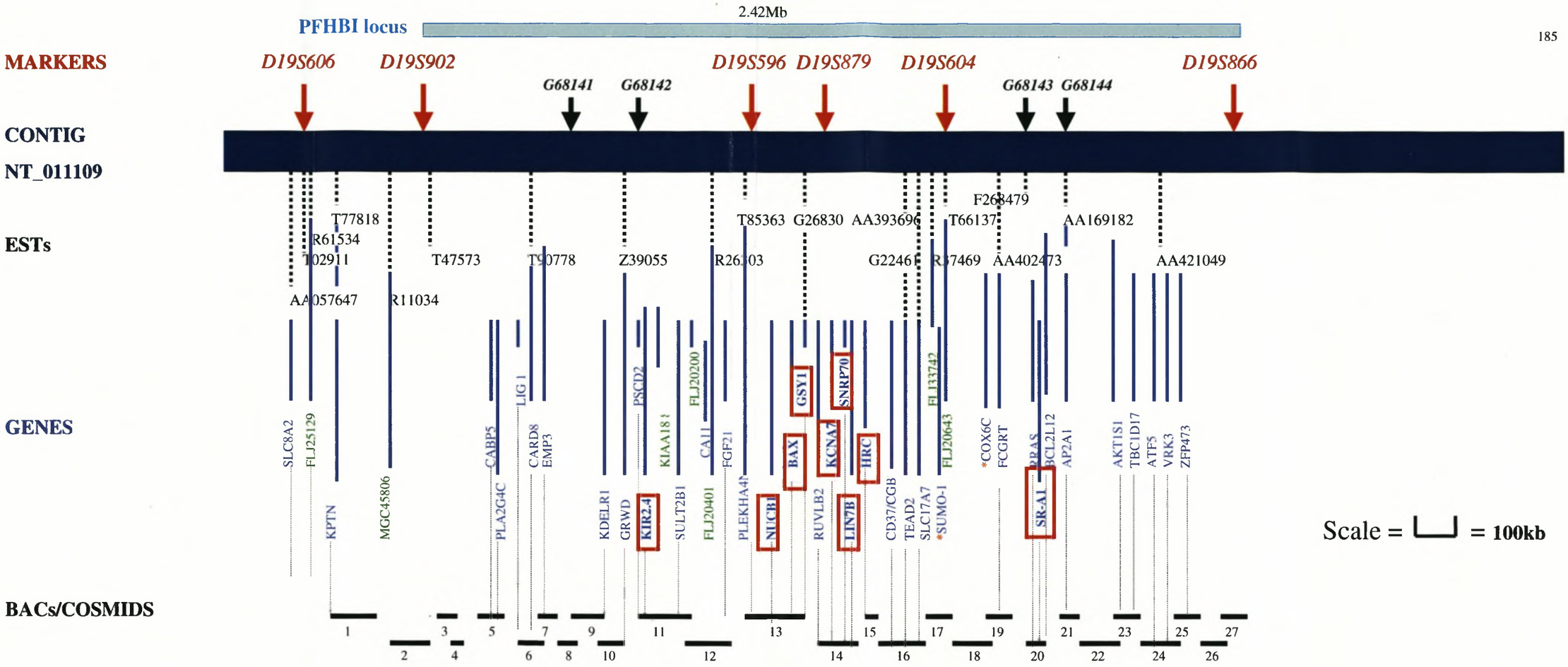


Figure 3.15 An integrated physical map of the PFHBI locus on contig NT_011109.

The positions of the published genetic markers are shown using red arrows. The novel CA repeat markers are shown using small black arrows. Fully annotated genes in blue , selected predicted genes in green , ESTs, the BAC or cosmid clones and novel genetic markers are shown in black on contig NT_011109.
From top to bottom: The position of the database and novel markers on contig NT_011109 which spans the PFHBI locus are shown. The ESTs are linked to the respective gene which in turn is linked to the respective BAC and cosmid clones numbered 1 to 27 by dotted black lines. Genes screened in this and parallel studies are boxed in red. Predicted proteins are denoted by FLJ, MGC and KIAA. The genbank acc numbers of the BACs and cosmids are:1, AC010331; 2, AC008745; 3, AC024582; 4, AC010330; 5, AC010458; 6, AC011466; 7, AC008392; 8, AC020955; 9, AC011514; 10, AC011527; 11, AC008403; 12, AC009002; 13, AC026803; 14, AC008687; 15, AC008891; 16, AC011450; 17, AC010524; 18, AC010643; 19, AC010619; 20.AC011495; 21, AC006942; 22, AC018766; 23, AC024079; 24, AC011452; 25, AC010642; 26, AC008655; 27, AC020909. The gene abbreviations are: *SLC8A2*, solute carrier family 8 (sodium independent inorganic phosphate cotransporter), member 2; *KPTN*, actin binding protein; *CABP5*, calcium binding protein 5; *PLA2G4C*, phospholipase A2, group IVC (cytosolic, calcium-independent); *CARD8*, caspase recruitment domain family member 8; *EMP3* epithelial membrane protein 3; *KDELR*, (Lys-Asp-Glu-Leu) endoplasmic reticulum protein retention receptor 1; *GRWD*, glutamate rich WD repeat protein; *PSCD2*, pleckstrin homology, sec 7 and coiled –coil domain 2; *KIR2.4*, potassium inwardly-rectifying channel, subfamily J, member 1; *SULT2B*, sulfotransferase family, cytosolic, 2B, member 1; *CA11* carbonic anhydrase X1; *FGF21*, fibroblast growth factor 21; *PLEKHA4*, pleckstrin homology domain containing, family A (phosphoinositide binding specific) member 4; *NUCB1*, nucleobindin 1; *BAX*, *BCL2*-associated **X** protein; *GYS1*, glycogen synthase 1; *RUVBL2* RuvB-like 2 *(E. coli); KCNA7*, potassium voltage-gated channel, shaker-related subfamily, member 7; *SNRP70*, small nuclear ribonucleoprotein 70kDa polypeptide (RNP antigen); *LIN-7B*, lin-7 homolog B (C. elegans); *HRC*, histidine richcalcium binding protein ; *CD37*,
CD37 antigen; *CGB7*, chorionic gonadotrophin beta polypeptide 7; *TEAD2*, TEA domain family member 2; *FCGRT*, Fc fragment of IgG, receptor, transporter, alpha ; *SLC17A7*, solute carrier family 17 (sodium independent inorganic phosphate; *RRAS*, related RAS viral oncogene homolog; *SR-A1*, serine arginine-rich pre-mRNA splicing factor; *ATF5*, activating transcriptor factor 5; *VRK3*, vaccinia related kinase 3; *ZNF473*, Zinc finger protein 473. *SUMO-1,small ubiquitin-related modifier-1 and *COX6C, cytochrome oxidase VIC, represent predicted genes identified from syntenic mouse ESTs.

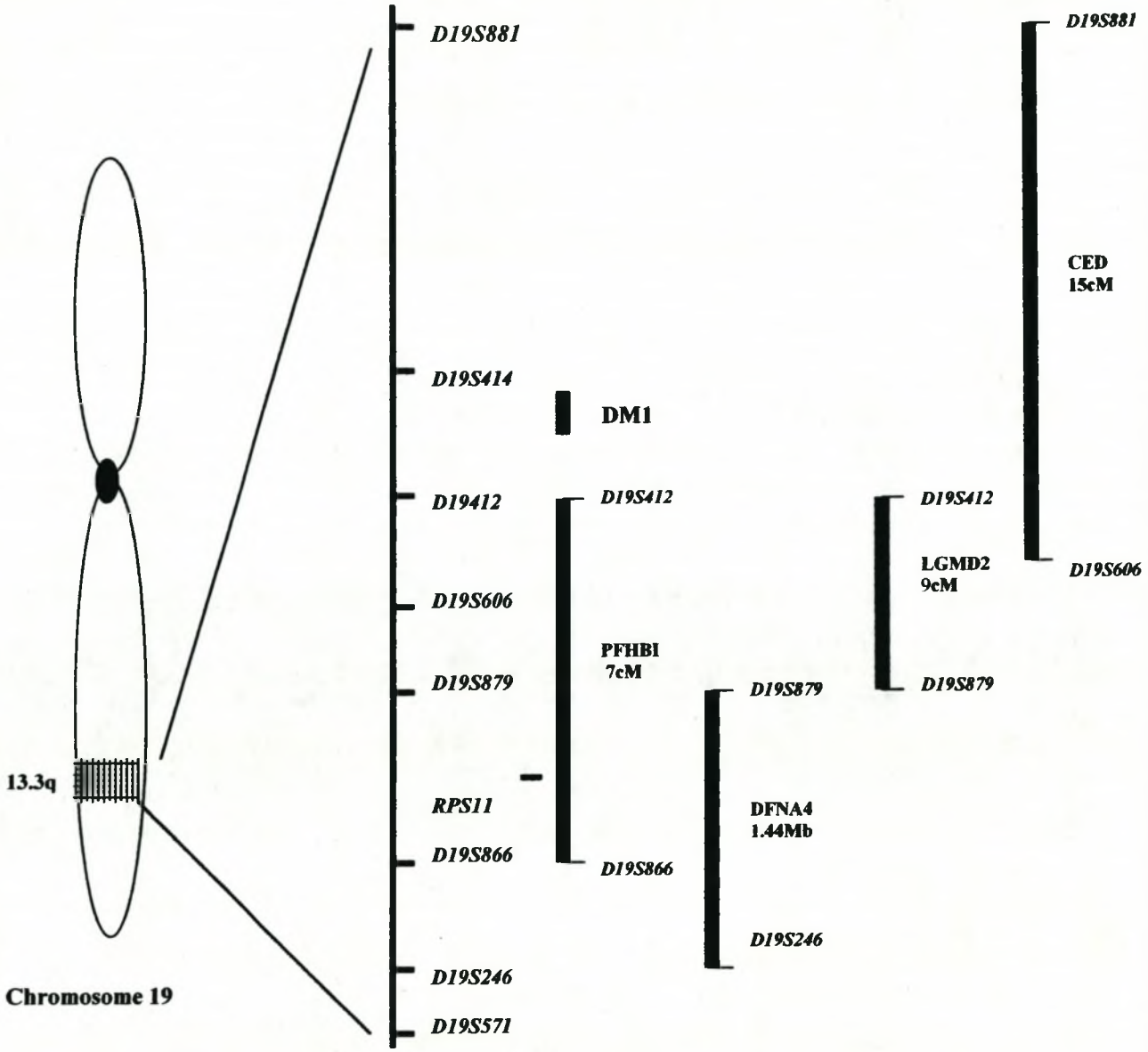


Figure 1.14 Schematic morbid map of chromosome 19q13.3.
Loci of different diseases mapping in the region of the PFHBI locus are indicated with a solid black vertical bar. The genetic map with markers mapping in the area is shown. RPS11, ribosomal protein S11; DM1, myotonic dystrophy I; PFHBI, progressive familial heart block I; LGMD2, lower girdle muscular dystrophy; DFNA4, dominant nonsyndromic deafness; CED, Camurati-Engelmann disease. cM, genetic distances of disease loci; Mb, physical distance in megabases.

flanked by genetic markers, *D19S881* and *D19S606* (fig. 1.14) (Ghadami *et al.*, 2000). The CED and the PFHBI loci overlap in the following way: the distal CED marker, *D19S606*, lies within, and the proximal *D19S881* marker lies centromeric to, the PFHBI locus (fig. 1.14). It was found that the transforming growth factor-beta-1 gene (*TGFB1*) maps to the same region of chromosome 19 and is a potential candidate gene for CED. Kinoshita *et al.*, (2000) screened *TGFB1* for mutations in seven unrelated Japanese families and two families of European descent and detected three different heterozygous missense mutations in exon 4, near the carboxy-terminus of the latency-associated peptide (LAP), in all nine families examined.

An autosomal dominant nonsyndromic deafness locus, DFNA4, was mapped to chromosome 19q13.3 by linkage studies and is flanked by genetic markers, *D19S414* and *D19S246* (fig. 1.14) (Chen *et al.*, 1995). The PFHBI locus lies within the DFNA4 locus and two genes mapping to this locus have been suggested as candidates for DFNA4, namely, *DMK*, and a gene-encoding mitochondrial ribosomal protein S12 (*MTRNR1*) (Johnson *et al.*, 1998; Shah *et al.*, 1998). Myotonic dystrophy is characterised by myotonia, in conjunction with progressive weakening and wasting of skeletal muscles, although central nervous system, ocular, cardiovascular and auditory manifestations have also been observed. Recently, the examination of a four-generation German kindred led to the identification of a 1.44Mb map segment in contig NT_011109 as being the most likely DFNA4 candidate region in chromosome 19q13.33, and placed the DFNA4 locus between the markers *D19S879* and *D19S246*. This region overlaps with the telomeric region of the PFHBI locus. More than fifty annotated and predicted genes were found in

the 1.44Mb region, of which several were promising DFNA4 candidates (Pusch *et al.*, 2004). To date, the disease-causing gene responsible for DFNA4 has not been identified.

1.4.5 Prioritisation of candidate genes for PFHBI

A number of genes have been characterised and placed at the PFHBI locus by LLNL, JGI, and NCBI mappers. These genes are all potential candidate genes for PFHBI by virtue of position, and prioritisation was a difficult task. For some of these genes, published information exists in the UNIGENE (NCBI) and LOCUSLINK (NCBI) databases regarding their tissue expression, function, exon and intron size and chromosomal location. The analysis of this information aids in prioritising genes for mutation screening, based on the appropriate function, tissue expression profile and developmental stage.

The HGP has produced hundreds of megabases of sequence; however, initially the sequence itself does not provide knowledge of gene structure, such as the number of exons or the functionally important domains. The initiative to have a 'rough draft' produced by 2001, shifted the efforts of the scientists involved to focus on the gene structure, regulatory regions and the translated products. A number of different gene prediction programmes, such as GeneScan (Burge and Karlin, 1997), the conserved domain database (CDD) (Marchler-Bauer *et al.*, 2002) and Pfam (Bateman *et al.*, 2002) were used to determine the structural organisation and the presence of important domains of predicted genes. After analysis, the predicted genes and their respective data were

submitted to Locuslink at NCBI databases and INFOGENE at Sanger Centre, UK (Solovyev and Salamov, 1999).

1.4.6 Mutation versus polymorphism

The human genome is not a static entity and is subject to a variety of heritable changes (mutations), which can include small changes, the so-called point mutations. These changes involve a loss, duplication or alteration of a single nucleotide of DNA (Cooper *et al.,* 1998). A DNA polymorphism is described as an allelic sequence variation where more than one allele occurs at a locus in a population with the frequency of the rarer allele being greater than 0.01 (Strachan and Read, 1996).

Cotton and Scriver (1998) examined the question 'when does a mutation have a deleterious effect on an individual?' A likely answer is that it might be dependent on the type and/or the position of the mutation. A base substitution at the DNA level, if synonymous, will not change the encoded protein, unless the base change is at an exon/intron junction, or in the regulatory region outside the exon, in which case, it may be pathogenic. These substitutions often occur at the third base pair position which often means that the altered codon specifies the same amino acid.

A nonsynonymous substitution can result in a nonsense mutation which is the replacement of a codon specifying an amino acid by a termination codon which is likely to result in a dramatic reduction in gene function. If, however, the nonsynonymous substitution is a missense mutation, one of two situations can arise, namely, a

conservative substitution which will result in replacement of an amino acid that is chemically similar and which therefore may not have any deleterious effect on protein function. The alternative being a nonconservative substitution in which one amino acid is replaced by another amino acid with a dissimilar side chain, which may be pathogenic, as it is more likely to affect protein function.

Protein-encoding genes show an enormous variation in the rate of nonsynonymous substitutions. At one extreme are proteins whose sequences are highly conserved, such as ubiquitin and calmodulin. For example, the ubiquitin proteins of *H. sapiens*, *M. musculus* and *D. melanogaster* show 100% sequence identity. Genes falling into this category have an extremely low rate of nonsynonymous codon substitution compared with other genes. The reason being that these proteins play such a crucial role and are, therefore, under selective pressure to conserve their sequence. At the other extreme, the fibrinopeptides are proteins which are evolving extremely rapidly and therefore appear not to have been under selective constraint.

Both a polymorphism and a mutation can be detected using various mutation screening techniques, such as heteroduplex analysis (Grompe, 1993), single-strand conformational polymorphism (SSCP) analysis (Orita *et al.*, 1989), denaturing gradient gel electrophoresis (Vijg and van Orsouw, 1999), the protein truncation test (Den Dunnen and Van Ommen, 1999) and direct sequencing of PCR products. These changes in the nucleotide sequences may cause frameshifts, splice sites or nonsense mutations, which in turn cause truncation of a protein product. The different methods usually test for

differences between a test and a standard sequence. These methods have shortcomings and have been critically analysed in reviews by Grompe (1993) and Prosser (1993). The final test to be performed is the functional analysis of the mutant gene, which will measure the biological activity of resultant protein confirming whether the mutation is phenotype-modifying.

Cotton and Scriver (1998) generated a list of criteria for the designation of mutations, which may modify the phenotype. The criteria are based on the particular types of mutation discussed earlier. They state that it is important to analyse the complete gene sequence and not specific exons of a gene because a "significant" mutation may be missed. In publications, it is therefore important to specify the regions analysed and the efficiency of the mutation detection method employed. Segregation analysis should be performed and reported. Reports should also mention whether or not a conserved amino acid is changed by the base substitution. A report on the prevalence of a mutation is important, because a mutation that occurs on less than 1% of the alleles in the population is, by definition, rare and may be disease-causing, whereas a variant that is found at higher frequency may be a "neutral" polymorphism. Reports of changes resembling polymorphisms should include the analysis of 100 chromosomes from the same ethnic group.

1.5 PFHBI CANDIDATE GENES

This section introduces each candidate gene selected by virtue of its position at the PFHBI locus for further analysis, as the possible PFHBI disease-causative gene. The organisation of each gene and the possible role of its protein product in cardiac function are discussed. The functional characteristics of each gene-encoding protein play a role in the selection of PFHBI candidate genes. In the present study, the genes selected for mutation screening were an apoptotic gene (*BAX*), two voltage gated K^+ channel genes (*KCNA7* and *KIR2.4*), a membrane-associated-guanylate kinase gene (*LIN-7B*) and a gene involved in glycogen metabolism (*GSYI*) and they are therefore described in more detail below.

1.5.1 *BAX*

BAX maps to the PFHBI disease locus on chromosome 19q13.3 (fig. 1.13) (Apte *et al.*, 1995). The BAX protein is encoded by six exons and demonstrates a complex pattern of alternative splicing with six alternatively spliced transcript variants encoding different isoforms (Oltvai *et al.*, 1993).

BAX induces cell death by acting on the mitochondrion, where it binds to the pore complex involved in the regulation of mitochondrial permeability (reviewed by Kroemer, 1998). Mitochondria play an important role in the signaling machinery of apoptotic cell death by releasing several apoptotic factors such as cytochrome c and procaspases. Furthermore, Ca^{2+} signals have been identified as one of the major signals to trigger the mitochondrion-dependent pathway of apoptotic cell death (Pacher *et al.*, 1999). The importance of apoptosis in cell growth and differentiation was only realised in the last

decade with the discovery of the cell death machinery in *C. elegans* and mammals. A delicate balance exists between cell proliferation and programmed cell death. Alterations in the apoptotic pathways can result in a number of diseases in humans. The activation of apoptotic pathways have been implicated in neurodegenerative diseases, such as Huntington's disease (Jellinger *et al.*, 2001), Alzheimer's disease (Sanchez *et al.*, 2001) and heart disease (Kumar *et al.*, 2002).

In the human heart, the morphogenesis of the CCS and the normal postnatal involution of the right ventricle are both mediated by apoptosis. However, the exact mechanisms by which these events are initiated or terminated remain poorly understood (James *et al.*, 1996). Although apoptosis is essential for normal development, excessive apoptosis may be triggered by pathological changes and result in the destruction of tissues and the development of heart disease in which a fatal arrhythmic event may occur. James *et al.* (1996) concluded the latter from histological examinations of the CCSs of a young woman and two of three brothers who had died of CHB. In two of the three hearts, the A-V node was absent, in the third heart, only fragments of the A-V node remained and in all three hearts, the sinus node was nearly destroyed by a non-inflammatory degeneration with no abnormal fibrosis or infiltrate. The histological abnormalities of the patients' CCS are best interpreted as resulting from apoptosis (James *et al.*, 1996). At present, the triggers for programmed cell death in disorders of the cardiac rhythm are not well understood, but it may be that a mutation in *BAX* makes it more effective to respond to inappropriate signals (Nerheim *et al.*, 2001).

1.5.2 Ion channels

Since a number of ion channels have been implicated in cardiac function and disease, the investigation of the two K^+ ion channel genes mapping to the PFHBI locus was vital. The one candidate was a voltage-gated shaker-related K^+ ion channel gene (*KCNA7)* (Chandy and Gutman, 1995; Kalman *et al.*, 1998; Bardien-Kruger *et al.*, 2002*)* and the other an internal K^+ rectifier gene (*KIR2.4*) (Töpert *et al.*, 2000).

K^+ ion channels belong to a family of voltage-gated membrane cation channels (Kv) that form aqueous pores through which K^+ can flow, and which are found in virtually all organisms (reviewed by Miller, 2000). These channels are responsible for a number of diverse functions including maintaining membrane potential, regulating cell volume and modulating electrical excitability in specialised cells. There are two broad classes of K^+ ion channels, as reflected in their primary sequence: a six-transmembrane helix voltage-gated channel (Kv) and a two transmembrane helix inward rectifier channel (Kir). The six-transmembrane segment of Kv, designated S1-S6, forms a module that controls the opening and closing of the pore. The first three segments (S1-S3) are inferred to be helical along most of their length and are located on the lipid-exposed periphery of the membrane-embedded complex. The fourth transmembrane segment (S4) is not lipid-exposed and is thought to form the voltage-sensing element of these channels. The P loop, the fifth and sixth transmembrane segments (S5 and S6) form the pore and selectivity filter (Pollard and Earnshaw, 2002). All Kv channels have a conserved domain, called T1, in the region near the amino-terminal side of the first transmembrane

helix (fig. 1.15). It has been shown that globular subunits of unknown function bind to the lower part of the T1 domain (reviewed by Miller, 2000).

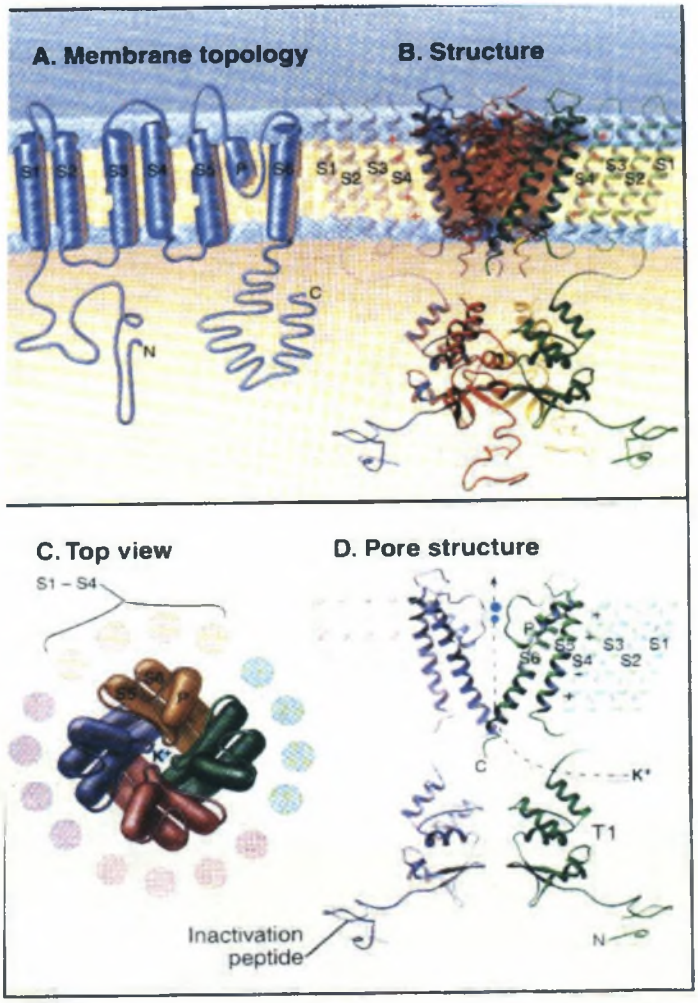


Figure 1.15 Voltage-gated potassium channels.

A, subunit transmembrane topology. Each of the four subunits has six transmembrane α-helices with a P loop between S5 and S6. B, ribbon diagram (for S5, P, S6), T1, and the N-terminal inactivation peptide. S1 to S4 are shown to be hypothetical helices lateral to the central pore-forming S5, P, S6. C, top view of the hypothetical arrangement of the six helices in each of the four subunits. D, side view of two subunits showing the central pore. Positively charged segment S4 senses the membrane potential. The T1 domain of K+ channels controls the assembly of four separate subunits into a functional channel (reproduced from Pollard and Earnshaw, 2002).

1.5.2.1 *KCNA7*

Kv channels in mammalian cells are encoded by an extended family of at least 19 genes (Chandy and Gutman, 1995). In humans, seven members of the subfamily, Kv1, have been identified and are designated Kv1.1-Kv1.7 (Chandy and Gutman, 1995; Kalman *et al.*, 1998). *KCNA7,* encoding the potassium channel Kv1.7, was mapped to chromosome 19q13.3 (fig. 1.13) (Kalman *et al.*, 1998) and at the initiation of this study, only the mouse homolog, *kcna7*, was characterised and had been mapped to mouse chromosome 7 (Kalman *et al.*, 1998). It was also shown that a region of mouse chromosome 7 was syntenic to the human *KCNA7* region and, during the present study, this data was used to determine the organisation of the human *KCNA7*, namely, that it is encoded by two exons separated by a 1742bp long intron (Bardien-Kruger *et al.*, 2002).

1.5.2.2 *KIR2.4*

There are seven subfamilies (*Kir1-Kir7*) involved in diverse functions, such as K^+ homeostasis, setting the membrane resting potential, modulating neural fusion rates, controlling pacemaker activity in the heart and, more recently, the Kir channels have been implicated in developmental processes (Doupnik *et al.*, 1995; Plaster *et al.*, 2001). *Kir* subfamilies have been found in the heart, skeletal muscle, endothelial cells and cells in the immune system (Kurachi *et al.*, 1999). Töpert *et al.*, (2000) cloned, characterised and mapped *KIR2.4* to chromosome 19q13.3 (fig. 1.13).

Recent studies reveal that the function and localisation of ion channels are regulated by interactions with members of the membrane-associated guanylate kinase (MAGUK)

protein family. It has been experimentally shown that there is a direct association of *Kir2.1*, *Kir2.2* and *Kir2.3* with the synapse-associated protein (SAP97), a MAGUK family member (Leonoudakis *et al.*, 2000). It was also shown that SAP97 might play an important role in the modulation of *Kv1.5* (Murata *et al.*, 2001). These findings suggest that *Kir* channels form part of a macromolecular signaling complex in many different tissues. Recently, it has been shown that mutations in a *Kir* gene, i.e., *Kir2.1*, cause dysmorphic features and cardiac arrhythmias in an inherited disorder, named Andersen's syndrome (Plaster *et al.*, 2001). Figure 1.16 shows the positions of the seven missense mutations and the two in frame deletions found in the different exons of *Kir2.1*. Since a *Kir* gene has been implicated in playing an important function in controlling cell excitability in skeletal muscle and heart (Plaster *et al.*, 2001), there is a strong possibility that mutations in other *Kir* genes may cause cardiac disturbances and are responsible for PFHBI.

1.5.3 *LIN-7B*

LIN-7B was mapped to the PFHBI locus between genetic markers *D19S902* and *D19S604* (fig. 1.13) (Strausberg *et al.*, 2002). It is encoded by five exons and is expressed in many tissues, including heart. The *LIN/VELI/MALS* family belongs to the *MAGUK* group of proteins, which contain multidomain modules, such as the PDZ domains (postsynaptic density-95, discs large, zonula occludens), which mediate protein-protein interactions. Using functional evidence, it has been shown that the PDZ domain of the MAGUK group interacts with the Shaker-subfamily K^+ channel, resulting in the clustering of the channels at the cell surface. In the case of Kv1.5, the interaction involves the COOH-terminal tail

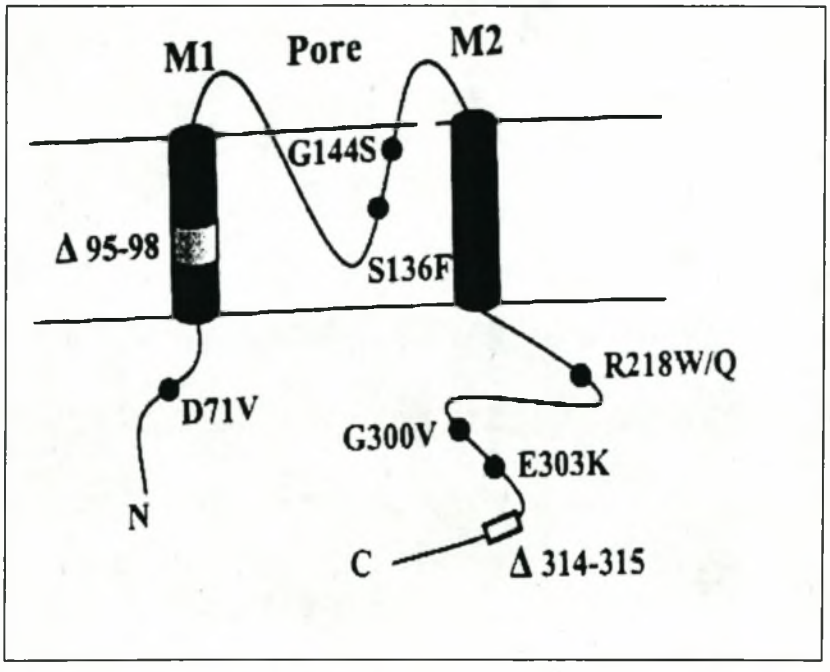


Figure 1.16 Schematic drawing of Kir2.1, an inward rectifying K^+ channel, outlining the position of the different disease-causing mutations in Andersen's syndrome.
M1 and M2, two transmembrane segments; D71V, S136F, G144S, R218W, R218Q, G300V and E303K, seven missense mutations; Δ95-98 and Δ314-15, in-frame deletions (Plaster *et al.*, 2001).

end of the K^+ channel with the PSD-95 protein of the MAGUK protein (Eldstrom *et al.*, 2002). This type of interaction was also shown to occur with Kir channels, where the PDZ domain of the human *Lin7/CASK* complex interacts with the COOH-terminal tail of Kir2.*3* in renal epithelia (Olsen *et al.*, 2002).

A study using *MALS*-1/2 double knockout mice demonstrated that the proteins containing the PDZ domains play a general role in recruiting receptors and enzymes at specific synaptic sites. A mechanism by which a PDZ interaction co-ordinates a one step basolateral membrane sorting programme was suggested (Misawa *et al.*, 2001). The role of *LIN-7B* in a signaling mechanism involving ion channels and its expression in the heart supports the hypothesis that *LIN-7B* interacts with *Kir* and, therefore, a mutation in one of these genes may be disease-causing, which makes both of them attractive candidates to screen for PFHBI disease-causative mutations.

1.5.4 *GSY1*

GSY1 cDNA has been cloned from human and rabbit skeletal muscle (Browner *et al.*, 1989) and from rat and human liver (Bai *et al.*, 1990). The muscle and liver isozymes are encoded by two different genes, muscle *GSY1,* which is located on chromosome 19q13.3 (position is indicated on fig. 1.13) (Lehto *et al.*, 1993) and *GSY2* on chromosome 12p12.2 (Nuttall *et al.*, 1994). *GSY1* is composed of 16 exons, which span approximately 27kb of DNA (Orho *et al.*, 1995) and specify a protein of 737 amino acids of which the NH_2- and COOH-termini contain nine phosphorylation sites, resulting in a very negatively charged molecule (Browner *et al.*, 1989).

GSY catalyses the rate-limiting step in glycogen synthesis and although glycogen occurs in abundance in the CCS, its role is unclear (James, 1970; Davies, 1971). However, it is known that glycolytic flux maintains membrane activity, reduces arrhythmia and improves the functional recovery of the ischaemic heart (King and Opie, 1998). *GSY1* has

been implicated as an important candidate gene for the genetic defect causing insulin resistance and non-insulin-dependent diabetes mellitus (NIDDM) (Orho *et al.*, 1995); therefore, a mutation in *GSY1* may affect the glycolytic levels, which may in turn affect the functioning of the heart and could possibly be PFHBI disease-causing.

Glycogen synthase kinase-3β (GSK-3 β) has been shown to inhibit glycogen synthesis through phosphorylation of GSY1. *GSK-3 β* is involved in the regulation of heart development via the *Wnt* signaling pathway (Hardt and Sadoshima, 2002). This could indicate that *GSY1* may play a role in the complex pathway of cardiac development. If this is so, it strengthens the case for *GSY1* as a candidate gene for PFHBI.

1.6 MANAGEMENT OF GENE SEARCHES

The search for the PFHBI gene is part of a team effort, a common situation in disease gene searches, examples of which include the identification of the genes causing FRDA (Campuzano *et al.*, 1996), RP13 (McKie *et al.*, 2001) and KWE (Findlay *et al.*, 1977). In the FRDA and RP13 gene searches, the disease-causative gene was found, whereas, the search for the KWE and the PFHBI genes are still ongoing.

The diseases RP13, KWE and PFHBI share a number of characteristics, for example, all three are autosomal dominantly inherited and affect South African families; many similar strategies are being used in the search for both the PFHBI and KWE genes. The strategies used and the results generated will be discussed below to give further insight into the

approaches used, the time span of a gene search, the pitfalls, the contributions of different team members and ideas of how to proceed with the search for the PFHBI gene.

1.6.1 Retinitis pigmentosa

Unlike KWE and PFHBI, RP13 is a genetically and clinically heterogeneous disorder characterised by progressive degeneration of the peripheral retina leading to night blindness and loss of visual field (Greenberg *et al.*, 1994).

In the early 1990's, the search for a gene causing autosomal dominant RP (adRP) was seriously pursued by a South African group under the supervision of Professor Jacquie Greenberg, based at the Human Genetics Department, University of Cape Town, South Africa. The RP13 locus for adRP was placed on chromosome 17p13.1 by linkage analysis (Greenberg *et al.*, 1994) in a large South African family of British ancestry. Since the recoverin (*RCV1*) gene mapped to chromosome 17p13.1 and was assumed to be involved in the pathogenesis of adRP, it was considered a prime candidate. However, mutational screening of all three exons of *RCV1* failed to reveal any mutations (Greenberg *et al.*, 1994) and at a later stage *RCV1* was excluded because of recombination events (Goliath *et al.*, 1995). Further studies, using the South African family, refined the RP13 locus to a region between markers *D17S1529* and *D17S831* flanking a distance of 3cM (Goliath *et al.*, 1995). A year later, an American group mapped the RP13 gene of another family of British descent living in the USA to chromosome 17p13.3 (Kojis *et al.*, 1996). At about the same time, an English group re-analysed the South African family data and refined the RP13 locus of their three-

generation family to chromosome 17p13.3 (Tarttelin *et al.*, 1996). Collaborations between the South African and the English groups were established and physical fine mapping of the region was actively pursued. The combined efforts resulted in the establishment of a YAC and PAC clone contig across the area. To further aid the search for the RP13 gene, an expression map of the region was constructed (McHale *et al.*, 2000).

As part of a doctoral study by Goliath (2000), a bioinformatic approach was used to establish the first sequence contig of the region. This study resulted in the reduction of the target region to 1cM and the identification of ten positional candidate genes, of which four were excluded from involvement in RP13 by mutation screening procedures. The fifth gene screened was a pre-mRNA splicing factor gene (*PRPC8*), which is composed of 42 exons and is ubiquitously expressed (McKie *et al.*, 2001).

The first 41 exons were excluded from harbouring RP13 disease-causing mutations, however, the final exon contained seven different mutations clustered in a 14-codon stretch of *PRPC8* in several unrelated RP families. It is therefore important to meticulously screen all the exons of PFHBI candidates as the disease-causing mutation may lie in the very last exon, as in the case of RP13 (McKie *et al.*, 2001).

1.6.2 Keratolytic winter erythema

Findlay *et al.*, (1977) first described KWE, which is also known as "Oudtshoorn skin disease", as an inherited dermatosis, which is traceable to certain nineteenth-century inhabitants of Oudtshoorn, Cape Province. Co-incidently, PFHBI (Brink and Torrington,

1977) was also first described in 1977. KWE is an autosomal dominant skin disorder characterised by erythema, hyperkeratosis and peeling of the palms and soles, especially during winter. This disorder occurs at a frequency of 1 in 7000 in the South African Afrikaans-speaking Caucasoid population and has been attributed to a founder effect (Starfield *et al.*, 1997). Interestingly, the belief about the frequency of PFHBI closely follows KWE, namely, 1 in 9000 in the same population group (Torrington *et al.*, 1986; Van der Merwe *et al.*, 1988).

KWE, like PFHBI, initially was thought to be unique to South Africa, but recently a single family in Germany was identified (Starfield *et al.*, 1997). Using a genome wide linkage analysis, the KWE locus was linked to marker *D8S550* on chromosome 8p22-23 with a maximum LOD score of 9.2 in a German family and 7.4 in five of the fourteen South African families (Starfield *et al.*, 1997). Eleven of fourteen South African KWE families studied, had an "ancestral" haplotype (Starfield *et al.*, 1997). New recombination events in the German family placed the KWE gene within a 10cM region, while analysis of recombination events in the South African families reduced the region to 6cM. An additional ancestral recombination event in the South African family further reduced the target region to a 1cM region flanked by markers *D8S550* and *D8S265* (Starfield *et al.*, 1997).

To identify positional candidate genes for KWE, a BAC contig spanning the KWE region, between markers *D8S550* and *D8S1695,* was constructed (fig. 1.17). The BAC clones spanning the KWE locus were identified by PCR, using primers of markers,

namely *D8S550*, *D8S265* and *D8S1695* at the KWE locus. All the BACs spanning the KWE locus were sequenced by Appel *et al.*, (2002), which resulted in the construction of a 634,404bp contig. A detailed physical and transcript map was constructed of the KWE locus by using various techniques, such as exon trapping, cDNA selection and sequence analysis, and twelve gene transcripts were identified (fig. 1.17). These transcripts were further characterised by determining the exon/intron structure and the expression profiles using reverse transcriptase-PCR (RT-PCR) with cDNAs from different tissues. Mutational analyses of each of the exons of the twelve transcripts were performed, but no potentially pathogenic mutations were found (Appel *et al.*, 2002). The search for the KWE gene is ongoing.

1.6.3 Progressive familial heart block I

The rationale for the PFHBI gene search and the earlier mapping data are discussed in detail in section 1.3.2.3.4. Saturation mapping by the group further narrowed the PFHBI locus to a 7cM region between markers *D19S412* and *D19S866* (fig. 1.9) (personal communication, de Jager, 1998). Due to a shortage of genetic markers within the PFHBI locus, the recombinant cosmid and BAC clones were probed for di- and tetranucleotide, such as $(CA)_n$, $(A_3T)_n$ and $(A_3G)_n$, repeat motifs for marker development. The novel markers, in turn, were used to further refine the map of the PFHBI locus. Over a period of time, a number of attractive candidate genes were selected for mutation screening by different members of the PFHBI group. One of the first genes selected for mutation screening was *HRC* (Hofmann *et al.*, 1991). The HRC protein is found in lumen of the

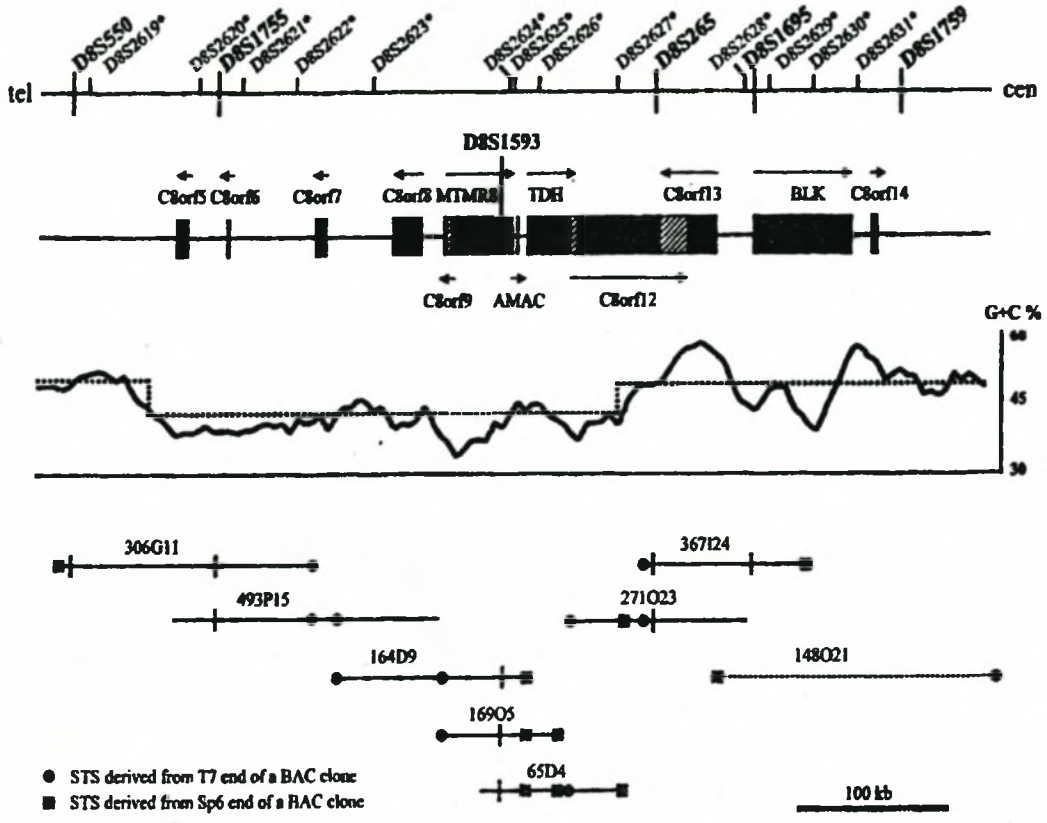


Figure 1.17 Physical and transcript map of the KWE critical region on human chromosome 8p22-p34.

Genetic markers in bold have been previously identified. The new microsatellite markers are marked with an * in the top map. The transcript map is shown underneath. The transcripts are depicted as filled boxes. EST locus *D8S1593* is part a novel myotubularin-related protein gene *MTMR8*. Partial overlap of the genes are shown in hatched boxes. The arrows show the direction of transcription. The G+C content of the genomic sequence is shown. The average G+C content is shown by the dotted line. The BAC clones, 306G11 to 148O21 were identified by PCR screening of a human BAC library. The BAC clones covering the region *D8S550* and *D8S1695* were completely sequenced. Within this sequence, markers *D8S550, D8S1755, D8S265* and *D8S1695* were identified. The contig was extended by partially sequenced BAC clone RP11-148O21 (Appel *et al.*, 2002).

sarcoplasmic reticulum and it has been proposed that sarcoplasmic proteins may play a role in Ca^{2+}-binding,therefore mutations in this gene could affect intracellular Ca^{2+} homeostasis.

As part of his study, Christoffels (1997) analysed the two triplet repeats, GAG and GAT, within exon 1 of *HRC* for abnormal expansions that might cause disease, since the phenomenon of anticipation was observed in the PFHBI family (Van der Merwe *et al.*, 1988). The $(GAT)_n$ repeat motif comprises 12 to 15 repeats and is tetra-allelic, whereas the $(GAG)_n$ repeat motif comprises 8 or 9 repeats and is biallelic in a normal Caucasion population. His investigation showed no expansions of either of these triplets in PFHBI-affected individuals. De Meeus and colleagues (1995) also investigated the members of the Lebanese family for repeat expansions of the two triplet repeats in exon 1 of *HRC* but no expansion was found. In addition, Christoffels (1997) developed a polymorphic $(CA)_n$ marker, *D19S1172*, which was genotyped in the PFHBI family and confirmed the telomeric limit of the PFHBI locus at *D19S866.*

In a more recent study, Makubalo (2000) screened the entire exon 1 sequence and the remaining exons of *HRC* (exons 2 to 6) using PCR-single-strand conformation polymorphism (PCR-SSCP) analysis. In addition, in 1998, a collaborative project with Dr Brian Black of the Cardiovascular Research Institute (University of California, San Francisco, USA) was formed which involved sequencing of the entire *HRC,* including the 5'and 3' UTR, exonic and intronic sequences of two PFHBI-affected individuals from

each of the pedigrees 1, 2 and 5 and two unaffected close relatives. However, no disease-causing mutations were found (personal communications, Black, 2001).

GSY1 was the next gene selected by the PFHBI group for mutation screening, based on the reasoning that although glycogen occurs in abundance in the CCS, its role is unclear (James, 1970; Davies, 1971). Five (4, 5, 11, 12 and 16) out of sixteen exons of *GSY1* were selected for mutation screening for PFHBI mutation screening by Makubalo (2000), using PCR-SSCP. These particular exons were selected because they contained domains which are conserved across the species. The proposal was that any sequence variation occurring in these regions might have a deleterious effect on the function of the protein. However, no mobility shifts were observed in any of the five exons examined. The remaining *GSY1* exons and the UTR regions were screened as part of the study described in this thesis.

In a more recent study by F. February (2002), all thirteen exons of *NUCB1* were screened for PFHBI disease-causing mutations by direct sequencing. The rationale for the selection of the human *NUCB1* gene for PFHBI mutation screening was that studies showed that lupus-prone mice secreted a murine *nucb1* protein that promotes the production of DNA-specific antibodies (Kanai *et al.*, 1993). These antibodies were also shown to be present in babies of mothers with systemic lupus erythromatosis (SLE) and one of the clinical symptoms of SLE is heart block in the absence of muscle failure (Lanham *et al.*, 1983). The screening of the exonic and flanking intronic sequences of all the exons of the *NUCB1* in PFHBI-affected individuals revealed no disease-causing mutations (F. February, 2002).

In 2002, another member of the PFHBI group screened the small nuclear ribonucleoprotein 70kD polypeptide (*SNRP70*) and serine arginine-rich pre-mRNA splicing factor SR-A1 (*SR-A1*) encoding genes (Du Plessis, 2004). The rationale was that defective mRNA splicing might result in disease (Faustino and Cooper, 2003), which implicates the involvement of splicing factor genes, since they are involved in the development of the CCS. Therefore, a mutation in genes involved in the splicing mechanism could possibly be PFHBI-causative. Although, at first sight, involvement of splicing factors would appear as an unlikely cause of a specific CCS defect, previous studies showed that a mutation in the *PRPC*8 gene, which is a splicing factor pre-mRNA, cause RP13 (for more detail on RP13, see section 1.6.1). In the case of *SNRP70,* no disease-causing mutations were found in PFHBI-affected individuals. In the case of *SR-A1*, both the SR (serine, arginine) and the DR (aspartic acid, arginine) domains in exon 6 were screened, but again no disease-causing mutations were found.

As part of Christoffel's (2001) doctoral studies, and as part of the PFHBI/SANBI collaboration, he identified 1184 fragment sequences of DNA clones spanning the PFHBI locus, which he screened using *in silico* methods, for microsatellite identification. In addition, he produced an integrated map of the PFHBI locus by assembling the 1184 sequenced fragments, mostly in draft form. To this, he mapped 119 STACK transcripts, 24 BodyMap transcripts, mouse ESTs and six RefSeq contigs. He identified seven cardiac positional candidates, which included actin-binding protein (KPTN), transcription factor (*TEAD2*) and a signal transduction protein, CD37, as possible disease-causative genes for PFHBI.

1.7 THE PRESENT STUDY

The aim of this study was to find the PFHBI-causative gene. The first step involved the reduction of the PFHBI locus using published markers. Further refinement of the locus involved, initially, the development of tetranucleotide $(A_3G)_n$ repeat motifs as markers, from cosmid clones obtained from LLNL, USA, and subsequently, the development of dinucleotide markers from publicly available chromosome 19 sequence data. The sequences from these clones were used to develop polymorphic markers, which were subsequently used in PCR-based genotyping analysis of the PFHBI families.

The second step used was the identification and prioritisation of the genes present at the PFHBI locus. The strategy involved searching the publicly available databases for ESTs, which were extended using two clustering programmes (STACK, SANBI and UNIGENE, NCBI). The significant clusters were retrieved and searched against protein databases with the aim of identifying novel candidate genes.

In order to accelerate the search for the PFHBI gene, an integrated map of the PFHBI locus was generated. All available information relevant to genetic and physical data, such as genetic markers, ESTs, clusters, genes, BAC cosmid clones and contigs spanning the region were integrated into a combined map of the target region using bioinformatic tools and wet-bench techniques. The map information was updated on a regular basis by performing BLAST homology searches using the physical data.

Since chromosome 19 is one of the most gene-rich chromosomes, it was necessary to prioritise both predicted and annotated genes. Candidates were selected for mutation screening based on their function and expression profile. In addition, the PFHBI locus was searched for a *Cx* and a G protein-encoding gene following the rationale discussed in sections 1.2.2.2 and 1.2.2.3. The whole genome of selected affected individuals was scanned for pathogenic a CTG repeat expansion which were largely done by collaborators (section 1.4.1.3).

1.8 CONTRIBUTIONS FROM THE OTHER PFHBI TEAM MEMBERS

During the course of the present study, I supervised 4 B.Sc Honours and 2 M.Sc projects: "The retrieval, clustering and protein homology searches of ESTs at the PFHBI locus" (F. February, B.Sc Hons 2000), "Development of $(CA)_n$ repeat markers in the region between genetic markers *D19S604* and *D19S866* at the PFHBI locus " (M. Du Plessis, B.Sc Hons 2001), " The search for tetra and triplet repeats within the PFHBI locus with a view to screening a strong candidate gene for PFHBI-causative mutations" (M. McCabe, B.Sc Hons 2002), "Optimisation of the PCR reactions using fluorescently labeled published genetic markers spanning the PFHBI locus" (A. Cloete, B.Sc Hons 2002), "Identification and mapping of expressed sequence tags, ESTs, with a view to identifying and to screening an attractive candidate gene for PFHBI " (F. February, 2002) and "The use of the mouse regions syntenic to the PFHBI region for identifying novel genes and the mutation screening of a candidate gene(s)" (M. Du Plessis, 2004).

In the present study, the Schalling group screened the whole genome of selected PFHBI individuals for a pathogenic CTG repeat expansion. Dr S. Bardien-Kruger was involved in the expression studies and the mouse/human homology sequence studies of *KCNA7*, together we designed primers which were used for mutation screening of affected and unaffected PFHBI individuals and I was principally responsible for the characterisation of the two SNPs of *KCNA7*.

CHAPTER 2

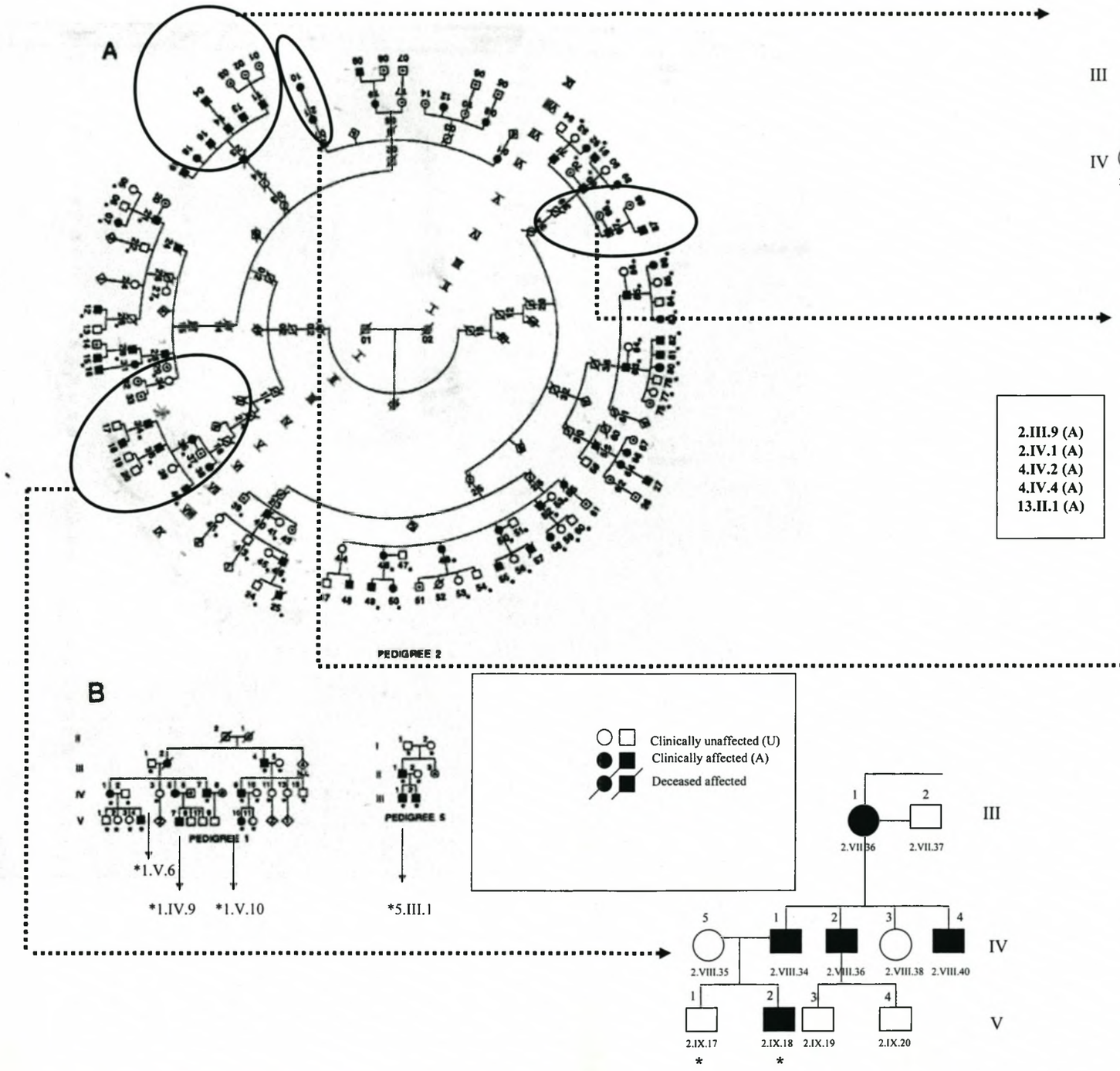
MATERIALS AND METHODS

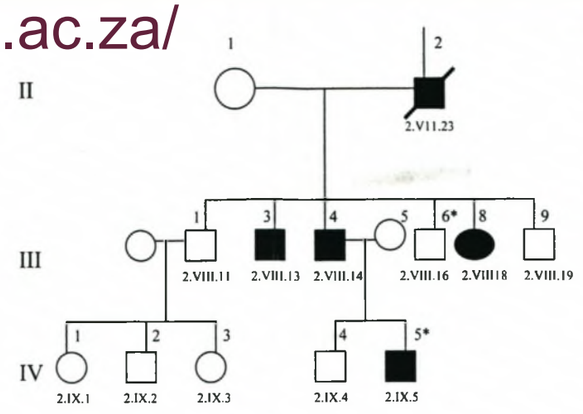
INDEX

2.1 SELECTION OF FAMILIES

Brink and Torrington (1977) first described a South African family, now designated pedigree 2, in which PFHBI segregates. Subsequently two smaller unrelated pedigrees, named pedigrees 1 and 5, were described. Pedigree 2 is the largest pedigree that can be traced back nine generations to a single male (I.01) and his wife (I.02). Individual I.01 was one of 10 siblings, the offspring of a Portuguese immigrant and his wife of French ancestory (see fig. 2.1A) (Torrington *et al.*, 1986; Brink *et al.*, 1994; Van der Merwe *et al.*, 1994). In this study, the different branches of pedigree 2 are referred to as kindreds (fig. 2.1). The smaller pedigrees (fig. 2.1B) display the same phenotype and the same haplotype as pedigree 2 (Brink *et al.*, 1995), and are therefore seen to represent a founder effect.

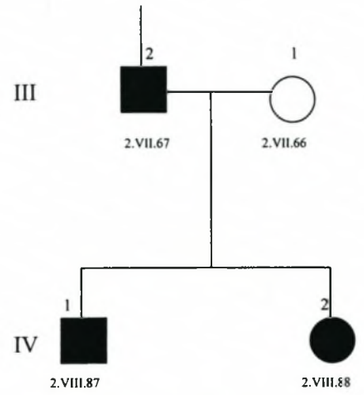
PFHBI-affected and unaffected family members were selected based on strict criteria with unequivocal clinical features described in section 2.2. In addition, affected individuals carried the disease-associated haplotype, except for individual 5.I.2 the disease was non-penetrant i.e., the disease-associated haplotype was present but the individual was not clinically affected (Brink, 1997). The control individuals were individuals that were married into the PFHBI pedigree, had a normal ECG and did not carry the disease-associated haplotype. Thus, for mutation screening studies, a panel comprising selected PFHBI-affected individuals from pedigrees 1, 2 and 5 and control individuals were used (fig. 2.1A and B). PFHBI-affected individuals from pedigree 2 were also investigated for the presence of a CTG repeat motif expansion (see the boxed samples in fig. 2.1A). The reason for selecting the CTG repeat motif for expansion is





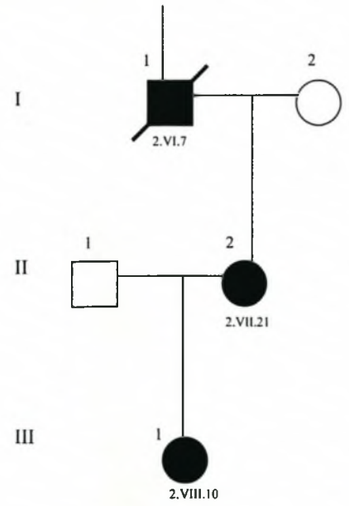
Pedigree 2 (kindred 9)

2.III.9 (A)
2.IV.1 (A)
4.IV.2 (A)
4.IV.4 (A)
13.II.1 (A)



Pedigree 2 (kindred 7)

Pedigree 2 (kindred 4)



Pedigree 2 (kindred 10)

Figure 2.1 PFHBI-affected family branches from pedrigrees 1, 2 (kindreds 4, 7 , 9 and 10) and pedigree 5.
The kindreds of pedigree 2 relevant to this study are enlarged and the nomenclature was changed. Subsequent to the previous study, newly affected individual 9.IV.5 was added to the study and changes to the clinical status of individuals were made, namely, individual 7.IV.2 is now affected and individuals 9.III.1 and 9.III.9 is classified as normal and . *, the DNA samples analysed for disease-causing mutations by PCR-SSCP analysis and direct sequencing. The boxed samples were investigated for CTG repeat expansions. Samples from pedigrees 1, 5, kindreds 4, 7, 9 and 10 were used for marker development and genetic fine mapping. In this study, the numbering system use for kindreds 4, 7, 9 and 10 of pedigree 2 is shown on the kindred drawing (Brink *et al.*, 1995).

addressed in subsequent chapters. Finally, individuals were selected from pedigrees 1, 2 and 5, which included siblings, children and married-in individuals, for genetic fine mapping studies (fig. 2.1A and B).

Additional DNA samples were obtained from 52 unaffected South African individuals of Afrikaner descent, with their consent, who were unrelated to the PFHBI families used in this study. These samples were used to determine the allelic frequency of the novel $(CA)_n$ markers and the frequency of single nucleotide polymorphisms (SNPs) of *KCNA7*.

2.2 CLINICAL EVALUATION

The clinical diagnosis of PFHBI was made by Dr P-L van der Merwe Department of Cardiology (Van der Merwe *et al.*, 1986; Van der Merwe *et al.*, 1988), Professor Brink, Department of Internal Medicine, Tygerberg Hospital (Brink *et al.*, 1994; Brink *et al.*, 1995) and personal physicians of referred patients, following strict criteria based on ECG-defined right RBBB, RBBB complicated by LAHB/LPHB or posterior hemiblock (bifascicular block) or CHB with a broad QRS-complex, in the absence of other cardiac conditions.

Patients with a pacemaker were accepted as of PFHBI-affected status, as it was not always possible to obtain a record of the cardiac rhythm that existed prior to implantation. Sinus bradycardia in isolation was considered normal and LAHB and LPHB in isolation were regarded as equivocal. Incomplete RBBB, indefinite

intraventricular conduction defect and first-degree heart block were regarded as normal in the absence of BB disease. The Minnesota code was adhered to in making ECG diagnosis (Higgins *et al.*, 1965).

2.3 BLOOD COLLECTION

This is a longstanding study, which involved the use of samples that had been banked, prior to the present study. The University of Stellenbosch's ethics committee approved the PFHBI project (Project 86/085). The blood samples were collected by Professors P-L Van der Merwe and P. Brink, the referred patient's physicians or Genetics Service nurses. All participants in the initial part of the study gave informed verbal consent and, in the case of minors, the consent of the parents was required. Written informed consent was required from new members participating in the study before any testing was done.

2.4 HUMAN GENOMIC DNA EXTRACTION FROM CELLS

DNA was extracted from either the nuclei of lymphocytes harvested from freshly collected blood samples or from Epstein-Barr virus transformed lymphocytes (modified Neitzel (1986) method), as described by Corfield *et al.*, (1993). The extracted DNA was air-dried and resuspended in 1ml of 1x TE (Appendix 1), mixed gently on a rotator (Voss of Maldon, Protea Dento-Medical Services, Cape Town) and quantitated using a spectrophotometer (Specronic 1201, Milton Roy). The concentration was adjusted to 0,2µg/ µl by dilution with 1x TE and was stored at 4°C.

2.5 PRIMER AND OLIGONUCLEOTIDE PROBE DESIGN AND SYNTHESIS

Oligonucleotide primers were designed for known STRs, for novel $(CA)_n$ STRs and $(A_3G)_n$ STRs from BACs and cosmid clones spanning the PFHBI region, for use in genotyping. Primers were also designed for mutation screening of exons and exon/intron boundaries of plausible candidate genes. Where possible, published primer sequences or database primers were used, but, when necessary, primer sequences were designed from published gene sequences with the assistance of the primer programme PRIMERFINDER (http://eatworms.swmed.edu/~tim/primerfinder). Primer sequences used in genotyping are given in table 2.1, those used for development of $(CA)_n$ dinucleotide and tetranucleotide STRs are given in table 2.2 and those used for the amplification of exons of candidate genes are given in table 2.3. All primers and the $(A_3G)_{10}$ oligonucleotide were synthesised at the DNA Laboratory based at the University of Cape Town (UCT), SA.

2.5.1 Primer design for known and novel STRs

The sequences of published primers of known STRs were retrieved from the Genbank database using Genbank accession numbers. Primers designed for novel $(A_3G)_n$ STRs from recombinant cosmid DNA are described in section 2.9. Primers designed for novel $(CA)_n$ STRs from sequenced BAC and cosmid clones are described in section 2.10.

2.5.2 Primer design for candidate genes

Primers for the amplification of exons of candidate genes were designed in such a way that approximately 30 nucleotides preceding and following the exon of interest were

included. This allowed examination of the intron/exon boundaries for the presence of disease-causing mutations. There were a number of large exons, for which primers were designed to amplify the exons in a series of overlapping products, such as both exons of *KCNA7* and exons 2 and 3 of *KIR2.4*. The design of *KCNA7* and *KIR2.4* primers is discussed in more detail in sections 2.14.4 and 2.14.5, respectively.

The sequence of each exon was retrieved from the Ensembl database (http://www.ensembl.org) and each exonic sequence was BLAST-searched (http://www.ncbi.nlm.nih.gov/BLAST) against the nonredundant database (http://www.ncbi.nlm.nih.gov/nr) to find its position on a BAC clone (Altschul *et al.*, 1990). Using the BAC clone sequence information, the flanking intronic sequences were retrieved. To confirm that the exonic sequence and the flanking intronic sequence were unique and on chromosome 19q13.3, a BLAST search was performed against the human genome database (http://www.ncbi.nlm.nih.gov/human genome) and the nucleic acid databases (http://www.ncbi.nlm.nih.gov/nr; http://www.ncbi.nlm.nih.gov/htgs).

2.6 PCR AMPLIFICATION AND SEQUENCING

2.6.1 Non-radioactive PCR

Non-radioactive PCR amplification was used in STR development which involved the amplification of cosmid inserts harbouring $(A_3G)_n$ repeat motifs and of database sequence of BACs and cosmids harbouring $(CA)_n$ repeat motifs. PCR amplification was also used to search DNA clones spanning the PFHBI locus for the presence of a *Cx (*see fig. 2.2)

Table 2.1 Database sequences of primers used to amplify published $(CA)_n$ repeat markers used in genotyping family members.

GENETIC MARKER	GENBANK Acc No	PRIMER 5'----3'	T_a °C	PRODUCT SIZE bp
D19S412	Z24042	F: tgagcgacagaatgagact	55	89-113
		R: acatcttactgaatgcttgc		
D19S606	Z53152	F: agggctgggacctcac	55	190-211
		R: ccaacacactgtctgcctt		
D19S902	Z53165	F: ccatcctaatgagggcaa	55	199-217
		R: gcaccagtgactgcctgt		
D19S596	Z52219	F: gaatccgagaggtggg	55	213-221
		R: gccagagccactgtgt		
D19S604	Z53032	F: tgtatgnaaagcacacgc	55	135-159
		R: atgtcatggagcatttgttg		
D19S879	Z52767	F: ctgagtgtgaatgaggcaac	55	251-278
		R aggccagagggggggactgattg		
D19S866	Z52072	F: catgagtttgactatgaagacg	55	183-217
		R: cactccagcctgggtaa		

The Genbank accession number (Acc No) for each genetic marker, the forward and reverse primer sequences (F and R), optimum annealing temperatures (Ta) and the expected amplicon size (bp) are indicated.

and a G protein-encoding gene. Finally, PCR amplification was used for mutation screening of the exons and exon/intron regions of the candidate genes, *GSY1*, *KCNA7*, *BAX*, *KIR2.4* and *LIN-7B*.

Table 2.2 Primer sequences for novel $(CA)_n$ markers generated during this study from recombinant BAC and cosmid sequences which were used in genotyping family members.

BAC/cosmid Acc No	Primer sequence 5' 3'	T_a ^0C	Product size bp
*AC008392	F: taggttggtcaggcttcgtg	56	248
	R: ctctgctttctgggttcaag		
AC008888	F: gtggatcacctgaggtcagg	59	226
	R: agctcctgttgcccaggctg		
*AC008403	F: gtgttgggattgcaggtgtg	59	249
	R: aggcaaaggatgcagcgaac		
AC010619	F: taacataaagggctgtcttc	50	169
	R: gtgaaaatagaacagcta		
*AC011495	F: ttgcagtgagctgagatcac	55	202
	R: tcgcatctagaacctattgc		
*AC018766	F: ctaatccactatggctggtg	55	171
	R: gaaatcttacaggggaaggg		
AC010624	F: gggtaaatgcttgactctcc	58	225
	R: agccagttcctgtccttatg		

The clone accession numbers (Acc No); the forward and reverse primer sequences (F and R); optimum annealing temperatures (T_a) and expected amplicon size (bp) are indicated; *, deposited in Genbank as $(CA)_n$ markers during this study (section 3.1.5 and table 3.1).

The general method for PCR amplification involved the preparation of a reaction mixture in a sterile 0.5ml microfuge tube which contained 150ng of each primer, 75μM each of dATP, dGTP, dTTP and dCTP (Promega Corp, Madison, Wisconsin, USA), 1.5mM $MgCl_2$, 0.5U of BIOTAQ™ DNA polymerase (Bioline, UK Ltd, UK) and 10x NH_4 reaction buffer (as supplied by manufacturer) (containing 160mM $(NH_4)_2SO_4$,

Table 2.3 Primer sequences used to PCR amplify exons and flanking intronic sequence of exons of candidate genes for mutation screening.

GENE	EXON	FORWARD PRIMER 5'—— 3'	REVERSE PRIMER 5'—— 3'	T_a °C	SIZE bp
***KCNA7**[a]					
	exon 1				
	1A	acacgtcggttcgcgggtcg	agtagagcacggcgtcgaag	55	200
	1B	ccgcgagtatttcttcgacc	tctcgggaaactcgaaaagc	55	200
	exon 2				
	2C	tggcacgcctgcgcgaggac	cgcccatctcctcccaaacc	55	200
	2D	atgtgtaaagaggagctggg	aagtagggaaggatagccac	55	200
	2E	gtccaagcaaggctatcttc	agaggaccacaccgatgaag	55	200
	2F	aatcttgggccagacgcttc	aattggagacaatgacgggc	55	200
	2G	aagatagtgggctctctgtg	tctagggaggtgtgaggtcc	55	200
***BAX**[b]					
	1	cgttcagcggggctctca	caggccggtaggaaggat	54	207
	2-3	cccctagaacccaagagtc	ggctgagagtcctgtgtcc	58	400
	4	tctcctgcaggatgattgc	tccccaggtcctcacagat	58	209
	5	caggcagtggggacaaggtt	gcggtggtgggggtgaggag	62	192
	6	cccctggccgagtcactgaa	aatgcccatgtcccccaatg	60	237
***KIR2.4**[c]					
	1	cactggactctctcagac	ccctgtatctcctttatatg	52	400
	2A	ctacatctcccacaatccc	gccttctggctcattgcctc	62	430
	2B	agctggccacgtgtgagaag	tgttgcccctgacgtttctg	62	450
	3A	gtgtagggttccagagcat	tggttccagaaagctcat	52	680
	3B	gcttgcctctgtggtcaacc	gtccccttccccgtgtatg	58	500
	3C	tccctgctgtctttggctcc	tacacggggaaggggacatc	62	691
***LIN-7B**[d]					
	1	caggtggtcggagcagtc	gggggctgggtatgtagg	60	220
	2	cgctctccttgcttctctgc	gacaagcaggaggaggagga	60	240
	3	gggagttgtagttttctcggcc	agggaggaaacagaaatgactgg	60	280
	4	gtaacccaggctaaatcgctgtc	ttgagtgaatgtcaaaagcaggg	60	340
	5	ggcaccaggctcactagcag	tggaagatccaggaagcactg	60	320

continued overleaf

Table 2.3 cont.

GENE	EXON	FORWARD PRIMER 5'---- 3'	REVERSE PRIMER 5'---- 3'	T_a °C	SIZE bp
#GSY1[e]					
	5'UTRA	ccgcctcagatccagaagtc	gagccagaagatctcgctcc	58	280
	5'UTRB	gatctcgatggttcctcata	ggccttcagctctgtcaatc	59	302
	5'UTRC	tttgaggcctccagctccag	agtagctggtgccgacgggaag	68	360
	1	gtcttgcaataggaagccgagcgt	catccactactgtccttctc	60	310
	2	cccagaactatggacctcg	gcaggctgtccacccgcttc	60	279
	3	gagcctccctatggtgcctc	gtcatgtagtttcaggtatgg	58	273
	6	accacaacattgatctgaatg	cctcctctcttaagacctag	57	213
	7	ggctttaccgtgccttgtg	ttgcttgccctccctgtcc	60	189
	8	ccacatgatcctggcctcc	ccggagcgagggctgctaac	61	167
	9	gcagctggtgtgtatgtgac	gagaacggagtaatgagagg	58	170
	10	cctctcattactccgttctc	ggctgccaggaactgctatc	58	172
	13	ttaagacggtaggcgatgag	aggagtcactgttctcaagc	58	207
	14	ctggagaccatgaagatctc	aggagctgctgatttgccag	57	270
	15	tggctctcacatcctctgcc	cctttagctcctggctaagc	60	163
	16B	ctgggcgaggagcgtaactaag	acttgatcgccccattcgcagg	68	438
	16C	ccaccaagttctggaaaccacgtg	gttgggaataagccaggttagggg	64	427
	16D	gatttagactctagagggaggag	catctcatctccggacacactc	64	378

T_a = annealing temperature; * = primer sequences generated during this study from database gene sequences; # = primer sequences obtained from literature; 5UTR = 5'-untranslated region; candidate genes: *KCNA7*, Potassium voltage-gated channel shaker related, member 7; *BAX*, BCL2-associated X protein; *KIR2.4*, Potassium inwardly-rectifying channel, sub-family J; *LIN-7B*, mammalian LIN-7 protein 2; *GSY1*, Glycogen synthase I; a = Bardien-Kruger *et al.*, 2002; b = Chou *et al.*, 1996; c = Genbank accession number NM_013348; d = Genbank accession number AF_311862; e = Ohro *et al.*, 1995.

670mM Tris-HCl, pH8.8 at 25°C, 0.1% Tween-20) and 5% formamide (Sigma Chemical Company, St Louis, Missouri, USA). To this, 200ng of human genomic /BAC/cosmid/radiation hybrid DNA was added and ddH_2O to a final volume of 50µl. A negative control was also prepared which contained all the components of the PCR reaction mix, except that the DNA was replaced by ddH_2O.

The reaction mixture was subsequently overlaid with 30µl mineral oil to prevent evaporation during amplification using the Gene-E-Thermal Cycler (Techne Ltd., Cambridge, UK). Each cycle typically comprised a 30 seconds (sec) denaturation at 94°C, incubation at the appropriate annealing temperature (Ta) for 30 sec and a 45 sec extension at 72°C. The annealing temperature selected was usually 3-5°C below the lowest melting temperature (T_m) of the primer set. The T_m for both the forward and the reverse primers were calculated as follows:

$$T_m = 4(G+C) + 2(A+T) \text{ (Innis and Gelfand, 1990)}$$

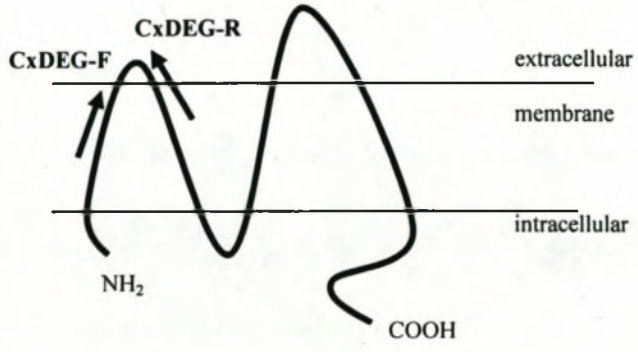
Tables 2.1 and 2.2 list the specific amplification parameters for all markers and table 2.3 lists the parameters used for the amplifications of exonic sequences of selected candidate genes used in this study. In the case of the *Cx* gene search, Cx degenerate primers were used (see fig. 2.2) (Coucke *et al.,* 2000) under the following PCR conditions: 1 min denaturation at 95°C, 1 min annealing at 55°C and 1 min extension at 72°C for 35 cycles.

2.6.2 Radioactive PCR

Radioactive PCR was used to genotype individuals from all three PFHBI pedigrees (see fig. 2.1) using published genetic markers, novel $(CA)_n$ repeat and tetranucleotide repeat markers. Radioactive PCR amplification involved the preparation of a mixture containing 60ng of each primer, 1µCi [α-^{32}P]dCTP (Amersham International, UK), 5µM unlabeled dCTP, 200µM of each of dATP, dTTP and dGTP, 1.5mM $MgCl_2$, 0.5U of BIOTAQ™ DNA polymerase (Bioline, UK Ltd, UK), 10x NH_4 reaction buffer (supplied by

manufacturer) and 5% formamide (Sigma Chemical Company, St Louis, Missouri, USA), which were added to 100ng of genomic DNA template in a 0.5ml microfuge tube and was made up to a final volume of 10µl using ddH_2O. The reaction mixture was subsequently overlaid with 15µl mineral oil to prevent evaporation during amplification.

(a)



(b) Primer CxDEG-F 5'-GGNAAGRTCTGGCTSACNGTNCTBTTCATHTTCCG-3'

Primer CxDEG-R 3'-CGDTAGAGGGTRYANGCGRWGACCCASGACGTC-5'

Figure 2.2 Generation of Cx degenerate primers (Coucke *et al.*, 2000).
(a) Schematic representation of the general structure of a connexin protein. The arrows indicate the location of the degenerate primers on a protein. (b) The nucleotide sequence of the forward (CxDEG-F) primer and the reverse (CxDEG-R) primer which was obtained from the alignment of six human connexins. The degenerate primers were designed and given as a gift by Coucke. The symbols in the degenerate primers represent: N = A/C/G/T, R = A/G, S = G/C, B = C/G/T, H = A/C/T, Y = C/T, W = A/T. A, adenine; C, cytosine; G, guanine; T, thymine.

After amplification, the samples were stored at room temperature in a perspex box until required for further analysis. In most cases, samples were stored for not longer than a month.

2.6.3 Direct Sequencing

2.6.3.1 Preparation of template

Direct sequencing of cosmids containing $(A_3G)_n$, which were identified by Southern blotting (Southern, 1975) described in section 2.9.1.3, was performed on both DNA strands. Firstly, 30µl of PCR amplification product was electrophoresed on an agarose gel and the DNA product was extracted using a GFX PCR DNA gel extraction kit (Amersham, UK) according to the manufacturer's protocol. The extraction of the DNA from the gel was performed as follows: After electrophoresis of the DNA on the agarose gel, the correctly sized DNA fragment was excised from the gel, by using a sterile razor blade for each sample to prevent contamination. After excision, the agarose band containing the DNA was weighed in a sterile microfuge tube. To the gel slice, 300µl capture buffer was added and the contents were mixed by vigorous vortexing, followed by incubation at 60°C in a waterbath until the agarose was completely dissolved. A new GFX column was washed with 250µl of capture buffer (buffered solution containing acetate and chaotrope) and left to stand for 1 min. A collection tube containing a GFX column was placed in a microfuge tube and was spun at 13,000rpm for 30 sec in a benchtop centrifuge (model Microfuge^R Lite, Beckmann Instruments, California, USA). The flow-through in the collection tube was discarded and capture buffer containing the dissolved agarose band was added to the column and spun at 13,000rpm for 30 sec in a

benchtop centrifuge (model Microfuge^R Lite, Beckmann Instruments, California, USA). The DNA was retained on a glass fibre matrix on the GFX column. The GFX column containing the DNA was washed with 500µl of wash buffer and was centrifuged at 13,000rpm for 30 sec in a centrifuge (model Microfuge^R Lite, Beckmann instruments, California, USA). Thereafter, the GFX column was placed in a sterile 1.5ml microfuge tube and to the column was added 50µl distilled sterile water, which was left to stand for 1 min at room temperature. In order to elute the DNA from the GFX column, it was spun at 13,000rpm for 1 min in a benchtop centrifuge (model Microfuge^R Lite, Beckmann Instruments, California, USA). The eluted DNA was used as a template for the sequencing reactions. The DNA was quantified spectrophotometrically.

2.6.3.2 Sequencing

In the initial part of the project, the inserts containing $(A_3G)_n$ repeat markers were sequenced using manual sequencing. Sequencing was performed using the fmol^{TM} sequencing kit (Promega, USA) according to the manufacturer's specifications. The sequencing reaction mixture contained 5µl DNA template, 5µl 5x sequencing buffer, 5% DMSO (2:3 dilution in ddH_2O) (Sigma, USA), 10pmol of primer, 0.5µl [α–^{32}P]dCTP (3000Ci/mmol) (Amersham International, UK), ddH_2O to 16µl and 5U of sequencing grade Taq polymerase (Promega, USA). Four microliters of the above cocktail mixture was added to four pre-chilled 0.5ml microfuge tubes, each containing 2µl aliquots of a cocktail composed of all four deoxynucleotides, as well as one specific dideoxynucleotide, namely, ddATP, ddTTP, ddGTP or ddCTP. Twenty microliters of mineral oil was carefully layered on top of the reaction mixture. Samples were then

placed into a pre-heated GeneE Thermal Cycler (Techne Ltd., Cambridge, UK) and kept at 95°C for 2 min before cycling was commenced. The same thermal cycling profile was used as that of the original PCR amplification. When the sequencing cycle was completed, 4µl sequencing stop solution (Promega, USA) was added to each tube. These samples were either stored at –20°C until required or denatured at 95°C in a heating block for 2 min and a 4µl aliquot immediately loaded on a pre-electrophoresed denaturing polyacrylamide gel (section 2.7.3).

2.6.3.3 Automated sequencing

In addition to manual sequencing, at a later stage of the project, PCR products of amplified exons of candidate genes were screened for mutations by direct sequencing, on an ABI Prism 3100 Genetic Analyser at the Core DNA Sequencing Facility, US. The templates used for sequencing were prepared as described in section 2.6.3.1 and the primers used for sequencing are described in section 2.6.1 (table 2.3).

2.7 GEL ELECTROPHORESIS

2.7.1 Agarose gel electrophoresis

Agarose gels were used for verification of the success of non-radioactive PCR amplifications, for preparative gels for sequencing template and for analysis of the purified BAC, cosmid and plasmid DNA. Five microlitres of either a PCR product or BAC/cosmid/plasmid DNA (0.1µg/µl) were mixed with 1µl of bromophenol blue loading dye (Appendix I). The samples were subsequently loaded and electrophoresed in 1.5% to

2% (depending on the size of the amplified PCR product) or 0.7% to 1% (depending on the size of cosmid and plasmid DNA) horizontal agarose gels (12cm x 5cm x 1cm), containing 1μg/ml EtBr and 1xTBE electrophoresis buffer (Appendix I). The samples were co-electrophoresed with a sample of bacteriophage lambda DNA digested with restriction enzyme *Pst*1 (λ *Pst1*) (Promega, USA) that was used as molecular size marker (Appendix I) or with a 100bp marker (Promega, USA). The gel was subsequently visualised on a longwave ultraviolet transilluminator (3UV^{tm} Transilluminator model LMS-26E) and a photograph was taken using an ITC camera (Berkenhoff and Drebes, Germany) and a Sony video graphic printer (UP-860CE).

2.7.2 Non-denaturing polyacrylamide gel electrophoresis

In order to screen exons of candidate genes for sequence variations, non-denaturing polyacrylamide gels were used for PCR single strand conformation polymorphism (SSCP) analysis according to methods modified from Orita *et al.*, (1989). For each panel of samples (fig. 2.1) analysed, four different non-denaturing polyacrylamide gel conditions were used, namely, 8% or 10% non-denaturing polyacrylamide gel, either with or without 10% glycerol, in order to maximise the likelihood of detecting sequence changes. In addition, a 0.35x MDE^{TM} (FMC Bioproducts, USA) gel was used for further analysis of samples where no SSCP mobility shifts were observed. Electrophoresis was performed at 4°C, since this prevented Ohmic heating (Liu and Sommer, 1994).

The casting of the gel

The gel was cast between 390mm x 290mm x 1mm glass plates and covalently bound onto Gelbond™ PAG polyester film (FMC Bioproducts, Maine, USA) during the gel polymerisation process (Appendix I). The fixing of the gel facilitated its subsequent handling during the staining procedure. The details of the casting of the gels and solutions are given in Appendix I. Once the gel was polymerised, the apparatus was dismantled in the following order, the casting boot, the sealing tape at the bottom of the

plates and lastly the comb were removed. The glass plate assembly was mounted vertically on the electrophoresis apparatus (Omeg Scientific, SA.), moved to the cold room (4°C) and 0.5xTBE buffer (Appendix I) was added to the reservoir tanks. The wells were washed with buffer to remove any gel residues, and the denatured PCR samples (section 2.6.1) were immediately loaded. Electrophoresis was carried out in 0.5xTBE buffer, at a constant power of 50W for 2-7 hours (hrs) at 4°C. For the separation of 100 to 200bp fragments, the fragments were electrophoresed until the bromophenol blue of the loading dye had migrated to the bottom of the gel. The separation of larger fragments (>200bp) was achieved by electrophoresing the xylene cyanol two-thirds of the gel length. Subsequently, the glass assembly was dismantled and the gel silver stained.

Silver staining

In order to visualise unlabeled amplified DNA fragments, silver staining was used. Following electrophoresis, the glass assembly was dismantled and the gel, still attached to the Gelbond™, was silver stained by immersion in solution B (Appendix I), on a

Labcon shaking platform (Labdesign Engineering, SA) for 10 min at room temperature, followed by a brief ddH_2O rinse and incubation in the solution C (Appendix I) on a Labcon shaking platform, for a minimum of 20 min or until the bands became visible. Thereafter, gels were again rinsed in ddH_2O and air-dried, photographed as a record and stored in a sealed plastic bag.

2.7.3 Denaturing gel electrophoresis

Following radioactive PCR amplification of STRs, or following manual sequencing reactions, samples were electrophoresed on denaturing polyacrylamide gels to separate alleles containing different numbers of repeat sequences or labeled fragments, respectively. The 0.5mm 5% polyacrylamide denaturing sequencing gel (Appendix I) was cast as described in section 2.7.2, but without the use of Gelbond™ film and using 0.5mm thin spacers and a sharks-tooth comb. Pre-electrophoresis was carried out in 0.5x TBE buffer at 1800V for 30 min prior to loading. The samples that needed to be electrophoresed were prepared in the following way: 6µl of sequencing stop solution (Appendix I) was added to each sample, which were then heat-denatured at 95°C for 5 min. Before loading, the wells of the gels were first washed out to remove any gel debris or air bubbles trapped in the wells. Thereafter, a sample volume of 3-5µl was loaded into the wells. The electrophoresis at 1800V continued for 2 hrs for genotyping STRs, however, for longer DNA sequencing reads, electrophoresis proceeded until the bromophenol blue dye had migrated more than two third of the way down the gel. At this point, the power was switched off and an additional 4µl aliquot of the same sample, heat–denatured, was loaded in a different lane for short sequence reads. Electrophoresis was

resumed at 1800V and allowed to proceed until the second bromophenol blue front had migrated more than two thirds of the way down the gel.

After electrophoretic separation was completed, the plates were dismantled, the gel was carefully lifted off the glass plates using 3MM Whatman paper (Whatman International Ltd, Madstone, England) cut to the same size as the gel plates, and thereafter covered with cling-film (GLADwrap^{TM}). Subsequently, the gel was dried for 1 hr on a Drygel Slabgel Dryer (model 1160, Hoeffer Scientific Instruments, San Francisco, USA), placed in a cassette and exposed to an X-ray film (Cronex-4, Protea Medical) overnight at room temperature.

2.7.4 X-ray development

The X-ray film (Cronex-4, Protea Medical) was developed, in the dark, by immersion in developing solution (Appendix I) for 4 min, stop solution (Appendix I) for 0.5 min and fixing solution (Appendix I) for 2 min. The X-ray film (Cronex-4, Protea Medical) was then rinsed off with running water and dried at room temperature.

2.7.5 Genotype analysis

The genotype scores were only assigned after agreement was reached between two independent individuals, blinded to the phenotypic status of the subject.

2.8 MUTATION SCREENING ANALYSIS

2.8.1 DNA sequence analysis

Each sample was sequenced in both the forward and reverse direction. The reverse complement of the reverse sequence of each sample was obtained using the BCM searchlauncher tool (http://www.hgsc.bcm.tcm.edu). Thereafter, the reverse complement of the forward sequence and the database sequence of each exon with flanking intronic sequences of each candidate gene were aligned using the CLUSTALW 1.81 programme (http://www.ebi.ac.uk/clustalw). All eight sequences were aligned against the database sequence retrieved from the NCBI database (http://www.ncbi.nlm.nih/nucleotide).

2.8.2 ClustalW 1.81 alignment of exons

The sequences obtained were aligned in the following order:-

1 = sequence of the exon from the database (http://www.ncbi.nlm.nih/nucleotide)

2 = forward sequence

3 = reverse complement sequence of the reverse sequence } Unaffected individual

4 = forward sequence

5 = reverse complement sequence of the reverse sequence } Affected individual from Pedigree 1

6 = forward sequence

7= reverse complement sequence of the reverse sequence } Affected individual from Pedigree 2

8 = forward sequence

9 = reverse complement of the reverse sequence } Affected individual from Pedigree 5

2.8.3 Submission of SNP to SNP database

A single nucleotide polymorphism (SNP) is defined as any polymorphic single nucleotide variation, which may or may not alter a restriction site (Vogel and Motulsky, 1986). In this project the PCR-SSCP technique and sequencing were used to detect SNPs. In addition, the frequencies of the polymorphic alleles of the SNP were analysed in 50 unaffected South African Afrikaner individuals. All the SNPs detected in this project were submitted to the SNP database (http://www.ncbi.nih.gov/SNP) (Wang *et al.*, 1998).

2.9 $(A_3G)_n$ MARKER DEVELOPMENT

As part of this study, the development of $(A_3G)_n$ repeat markers was initially pursued and subsequently $(CA)_n$ repeat markers were developed from available chromosome 19 sequences from public databases.

Twenty five Lawrist-16 cosmids containing fragments of human chromosome 19q13.3 (De Jong *et al.*, 1989), spaced at 500Kb intervals across the chromosome 19q13.3 region, were obtained from LLNL, California, USA. These cosmids were propagated in the DH5α strain of *E. coli*. In order to develop the $(A_3G)_n$ repeat markers, the approach outlined in fig. 2.3 was followed.

2.9.1 The identification of $(A_3G)_n$ repeat-containing recombinant cosmids (fig. 2.3A)

2.9.1.1 Cosmid culture preparation, isolation and purification of cosmid DNA

The recombinant Lawrist-16 cosmid cultures were prepared using the detailed method described in appendix I. The QIAGEN™ commercial DNA purification kit (QIAGEN, Germany) was used to isolate cosmid DNA from bacterial genomic DNA. Absorbance was read on a spectrophotometer at 260nm to determine the DNA concentration (µg/ml) using the formula: OD_{260} x 50 x Dilution factor ($1OD_{260} = 50\mu g$). The DNA was analysed on a 0.7% agarose gel containing 0.01% EtBr and 1xTBE to check its integrity (section 2.7.1). The cosmid DNA could now be utilised for the preparation of dot blots (section 2.9.1.2), Southern blots (section 2.9.1.3) or sub-cloning (section 2.9.4).

2.9.1.2 Preparation of dot blots

Dot blots were used for the initial screening of cosmids to identify those cosmids containing $(A_3G)_n$ repeats and again later to verify the presence of $(A_3G)_n$ repeats in the inserts. The denaturation of the cosmid DNA was performed in the following steps: to each 200ng of cosmid DNA an equal volume of 0.4N NaOH was added and the mixture was incubated at room temperature for 5 min. The mixture was snap-cooled on ice for 2 min and spotted onto the previously marked positions on a Hybond™ N^+ filter (Amersham, UK), which had been placed on 3MM-chromatography paper (Whatman, International Ltd, Madstone, England). Fifty nanograms of human genomic DNA were spotted onto the filter as a positive control and ddH_2O as a negative control. The filter was air–dried for 15 min and thereafter baked for 2 hrs at 80°C in a vacuum oven (Towson and Mecer Ltd, England). Subsequently, the filter was either re-hydrated in 2x

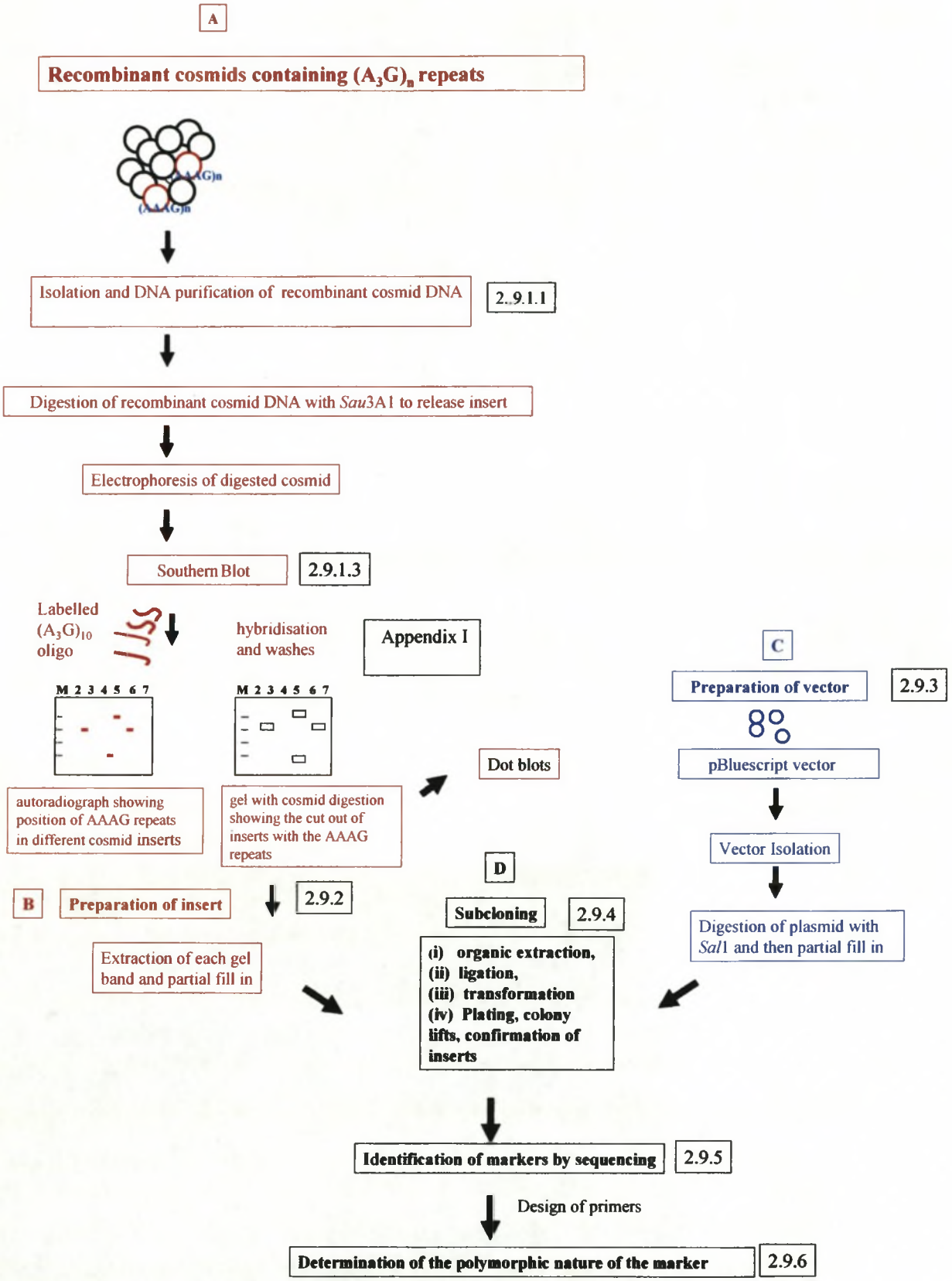


Figure 2.3 Schematic representation of the development of $(A_3G)_n$ markers
A, describes the steps taken to identify $(A_3G)_n$ repeats in recombinant cosmid clones. B, describes the steps involved in the preparation of the insert from (A_3G)n -containing clones. C, describes the preparation of the pBuescript vector. D, steps involved in the subcloning of vector and (A_3G)n-containing insert, sequencing of the insert and the determination of the polymorphic nature of the insert.

SSC (Appendix I) prior to the pre-hybridisation step of the screening process (Appendix I) or covered with clingfilm (GLADwrap^{TM}) and stored at 4°C until required.

In order to identify the $(A_3G)_n$ repeat containing cosmids to be used for subcloning, the Hybond^{TM} N^+ membrane was probed with an end labeled γ–^{32}P[dATP]-$(A_3G)_{10}$ oligonucleotide. A detailed description of the probe preparation, pre-hybridisation and post-hybridistion protocols are presented in Appendix I.

2.9.1.3 Cosmid DNA digestion and Southern blot preparation

The digestion reaction, which contained 30µg cosmid DNA, 90U *Sau*3A1 (Promega, USA), 15µl 10x buffer as specified by the manufacturer (Promega,USA), and ddH_2O to 150µl, was incubated at 37°C for 3 hrs, gel electrophoresed and Southern blotted (see fig. 2.3A) (Appendix I). Thereafter, the Hybond^{TM} N^+ membrane was either re-hydrated in 2x SSC (Appendix I) in preparation for probing with a radioactively labeled oligonucleotide (Appendix I) or stored at 4°C.

In order to identify the $(A_3G)_n$ repeat containing fragments to be used for subcloning, the Hybond^{TM} N^+ membrane was probed with an end labeled γ–^{32}P[dATP]-$(A_3G)_{10}$ oligonucleotide (Appendix I).

2.9.2 Insert preparation (fig. 2.3B)

For the purposes of subcloning, 30µg of cosmid DNA containing the $(A_3G)_n$ repeat were digested with 90U of *Sau*3A1 (Amersham, UK) according to the manufacturer's protocol

for 3 hrs and electrophoresed on 0.7% agarose gel (20cm x 20cm x 3cm) (Appendix I) at 30V overnight alongside λ *Pst*1 (Promega, USA). The *Sau*3A1 fragments were excised from the gel using a sterile blade and then the DNA was extracted from the gel slice using a GFX kit (Amersham, UK) according to the manufacturer's protocol (as described in section 2.6.3.1) and resuspended in 40µl ddH_2O.

2.9.3 Vector preparation (fig. 2.3C)

Plasmid vector pBluescript SK (Stratagene, USA) (fig. 2.4) cultures were prepared in a similar manner to the cosmid, as described in Appendix I except LB/ampicillin (Amp) (50µg/ml) selective medium was used (fig. 2.3B). The plasmid vector, pBluescript SK, molecular weight 2.9kb, was linearised by digesting with *Sal*1 restriction enzyme (Promega, USA). The digestion reaction mixture contained 10µg vector DNA, 4µl (100mM) spermidine, 200U *Sal*1 (Promega, USA), 10µl 10x buffer as specified by the manufacturer (Promega, USA) and ddH_2O to a final volume of 100µl. The reaction was incubated at 37°C for 3 hrs. Thereafter, linearised vector DNA was ethanol precipitated (Appendix I).

Partial fill-in reactions for insert and vector

In order to prevent self-ligation of the insert molecules and to render the ends compatible in the vector, dGTP and dATP were added to the insert in the presence of the Klenow fragment of DNA polymerase (Promega, USA). The fill-in reaction mixture consisted of 40µl (10ng) of insert DNA, 0.5µl dGTP (10mM), 0.5µl dATP (10mM), 2µl (10U) Klenow fragment of DNA polymerase (Promega, USA), 5µl 10x Klenow buffer

(Promega, USA) and ddH_2O made up to a final volume of 50µl. Incubation was carried out at 37°C for 2 hrs, after which 1µl (0.5M) EDTA was added to terminate each reaction and the reaction tube stored at 4°C.

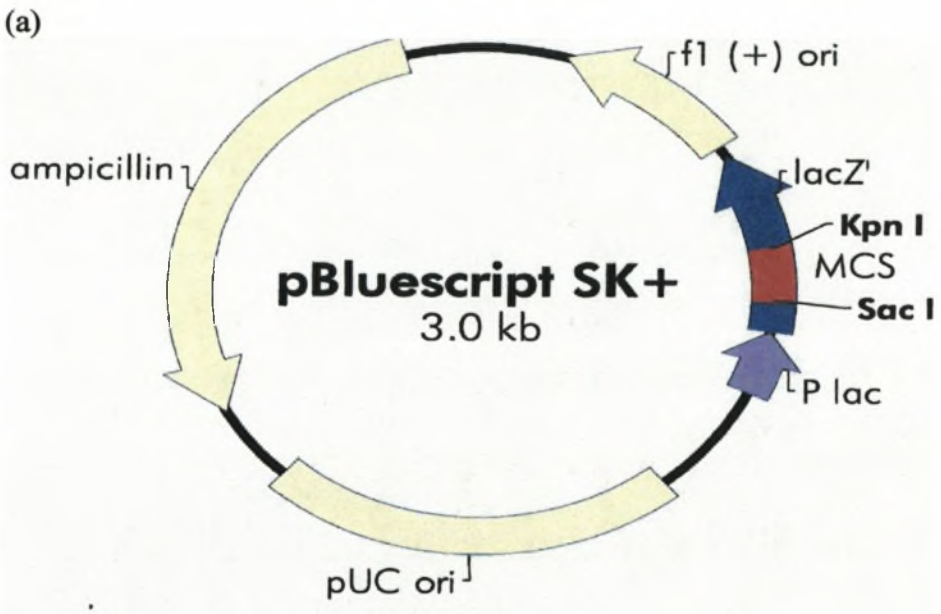


Figure 2.4 Diagram and restriction map of pBluescript SK vector.
(a) f1 (+) ori, f1 origin; *lacZ*, lactose Z gene ; MCS, multiple cloning site found within the *lac* Z gene showing enzyme restriction sites, *Kpn1* and *Sac1*; P lac, promoter of the lactose operon; pUC ori, pUC origin (b) The sequence of MCS showing the position of the different primer binding sites of M13, T3, KS and T7 and the different enzyme restriction sites found in MCS. (http://www.stratagene.com/vectors).

2.9.4 Subcloning of $(A_3G)_n$ repeat fragments (fig. 2.3D)

For subcloning, the partially filled insert and vector were purified using the organic extraction (Appendix I). The ligation of the vector and the insert, the transformation of

the *E. coli* XL1-blue competent using the ligation mixture and colony lifts of the recombinant clones are described in detail in Appendix I.

2.9.5 Identification and characterisation of $(A_3G)_n$ repeat motif (fig. 2.3D)

Recombinant plasmid DNA was extracted (section 2.9.4) and digested with *Kpn*I and *Sac*I at the vector restriction enzyme sites on either side of the multiple cloning site (MCS) (fig. 2.4), to verify that the positive colonies picked from the plates contained cloned insert (Appendix I). Following the verification of the presence of the insert, the corresponding recombinant plasmid clone were sequenced (section 2.6.3) using vector primers T3 and T7 (fig. 2.4).

2.9.6 Determination of the polymorphic nature of marker

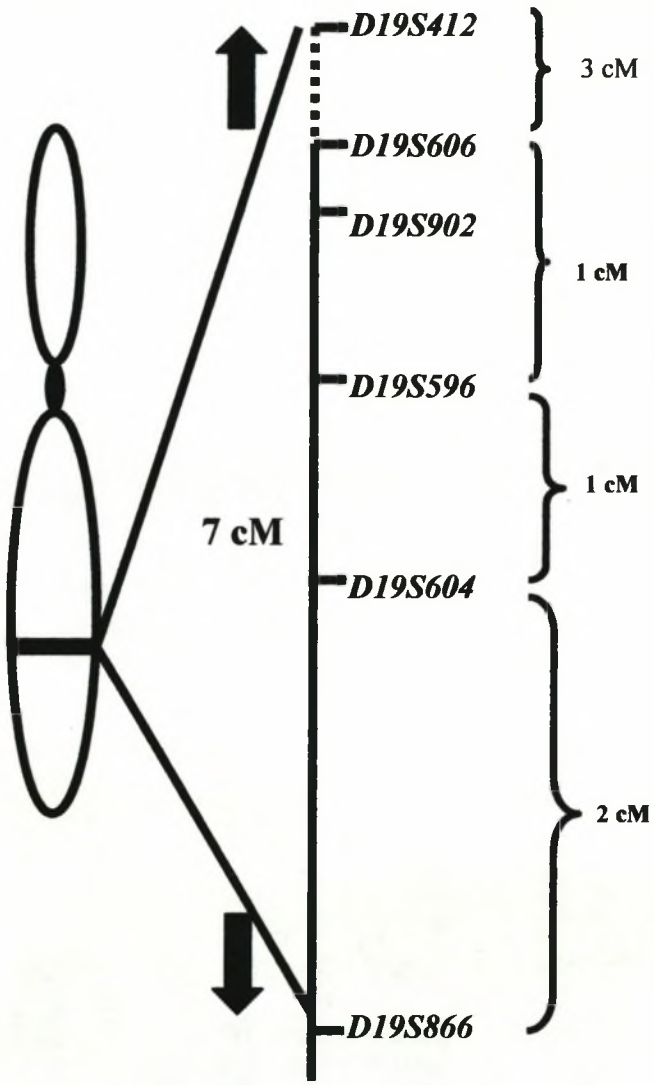
The polymorphic nature of each marker was determined by PCR amplification (as described in section 2.6.2) using DNA of 20-25 unrelated individuals of South African Afrikaner population. The PCR products were electrophoresed on a 6% denaturing gel as described in section 2.7.3 and the gel was exposed to X-ray film (section 2.7.4). Generally two alleles, one inherited from each parent per individual were observed on the autoradiograph. The alleles were numbered in order of decreasing molecular weight and the frequency of each allele was calculated as a fraction for a South African Afrikaner population group. DNA samples from selected individuals from kindreds 4 and 9 from pedigree 2 were also genotyped simultaneously, in order to fine map the PFHBI locus.

2.10 $(CA)_n$ MARKER DEVELOPMENT

2.10.1 Sequence acquisition

The availability of draft and final sequence from the Human Genome Project (HGP) has prompted the use of *in-silico* methods for the identification of STRs (Benson, 1999). During the course of the study, $(CA)_n$ genetic markers from the PFHBI locus on chromosome 19q13.3 had been utilised and the necessity arose to identify novel markers to refine the existing breakpoints in the critical area.

In-silico identification of tandem repeats became feasible with the release of the complete draft and final sequence for chromosome 19 in June 2000 (http://www.jgi.doe.gov) in the form of sequenced recombinant BACs and cosmids in the Joint Genome Institute (JGI) (fig. 2.5). Using both an *in-silico* approach and experimental techniques a number of CA/GT repeat markers were developed (fig. 2.6) using these sequences. A total of 1184 sequences spanning a 4Mb region of the PFHBI locus on chromosome 19q13.3 were downloaded from the JGI ftp site (ftp://sawdoff.llnl.gov/pub/JGI_data/Human/Ch19) as part of the PFHBI/SANBI collaboration (Christoffels, 2001). Sequences deposited into the publicly available databases are contaminated with *E. coli* sequences (Hillier *et al.*, 1996) and these contaminating sequences were removed by masking the *E. coli* sequences by using the programme Repeatmasker (http://ftp.genome.washington.edu/cgi-bin/RepeatMasker) prior to screening for tandem repeats (Christoffels, 2001).



LLNL Clone ID	UNIVERSAL Clone name	GENBANK Acc No.
BC821616	CITB-E1_3001H11	AC010331
F24003	LLNL-F_198H7	AC008985
R30005	LLNL-R_261D9	AC010519
BC782556	CITB-E1_257L23	AC008745
BC858854	CITB-E1_3098H1	AC024582
BC815354	CITB-E1_2657C13	AC010330
BC694629	CITB-H_2265M8	AC010458
BC324323	CIT-HSPC_453G23	AC011466
BC242886	CIT-HSPC_241F20	AC008392
R33773	LLNL-R_300F9	AC020955
F16353	LLNL-F_119C1	AC011514
R26730	LLNL-R_227C10	AC011527
BC255070	CIT-HSPC_273B12	AC008403
R29279	LLNL-R_253H3	
BC677569	CITB-H1_2221F12	AC008888
R31763	LLNL-R_279G3	AC009002
BC808483	CITB-E1_2639E6	AC026803
BC52309	CIT978SKB_60B18	AC008687
BC679592	CITB-H1_2226J19	AC008891
BC266129	CIT-HSPC_30107	AC011450
R31681	LLNL-R_278H5	AC010524
R29295	LLNL-R_254A7	AC010643
BC878087	CITB-E_3148I10	AC010619
BC42053	CITB78SKB_33G10	AC011495
R31181	LLNL-R_273F9	AC006942
F23669	LLNL-F_195D9	AC018766
R28785	LLNL-R_248GI	AC024079
BC275645	CIT-HSPC_326K19	AC011452
BC641056	CITB-H1_2126E3	AC010624
BC61330	CIT978SKB_191K22	AC008655
R28901	CITB-E_254M3	AC020909

Figure 2.5 Genetic map of the PFHBI locus showing the positions of recombinant cosmid and BAC clones (February 5, 2003).

From left to right: the genetic map of the PFHBI locus, LLNL clone ID, Universal clone ID and the Genbank accession number for each cosmid and BAC clones present at the PFHBI locus is shown. The cosmid clones are indicated in green and the BAC clones are indicated in blue.

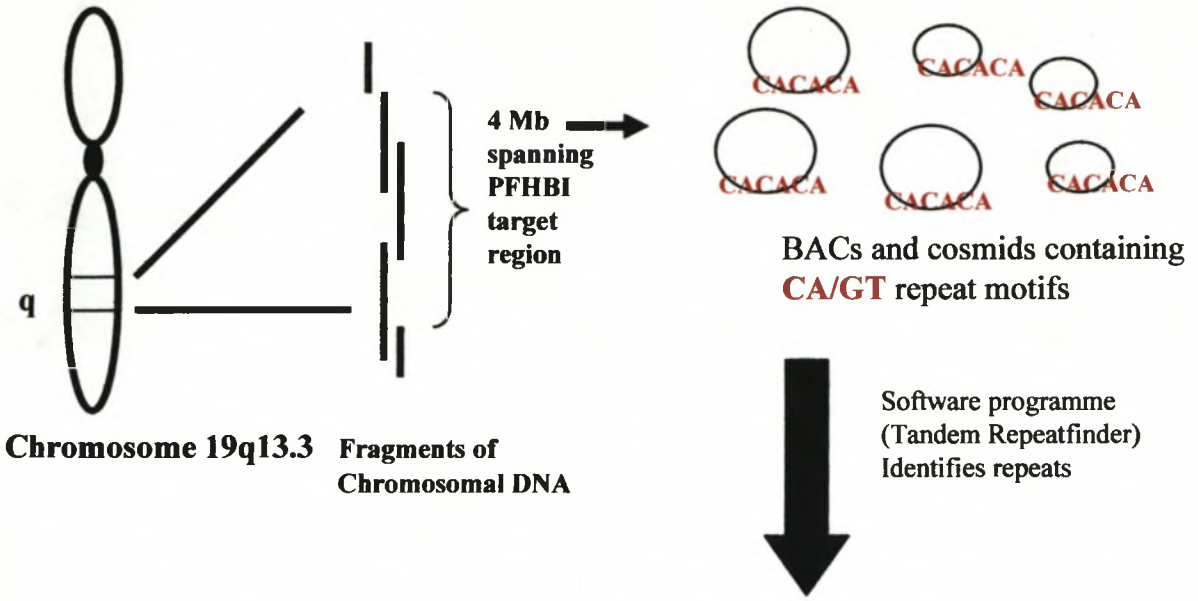
2.10.2 Screening cosmid and BAC sequences for tandem repeats

Tandem Repeatfinder (http://c3biomath.mssm.edu/trf.basic.submit.html) (Benson *et al.*, 1999) is a programme designed to detect long stretches of twelve or more tandem repeats. Each of the 32 cosmid and BAC sequences spanning the PFHBI locus were first masked for contaminating vector, simple repeats, mitochondrial and ribosomal regions and *E. coli* sequences after which it was subjected to Tandem Repeatfinder in order to detect all $(CA)_n$ repeats (fig. 2.6) (Christoffels, 2001). All the $(CA)_n$ repeat-containing cosmid and BAC sequences were downloaded and placed at http://ziggy.sanbi.ac.za/alan/TandemRepeats.html. (personal communication, Christoffels, 2001). As part of the present study, 27 BAC and cosmid sequences containing $(CA)_n$ repeats spanning the PFHBI locus were retrieved and investigated for marker development.

2.10.3 Protocol for marker development

2.10.3.1 Primer generation

Using the programme Primerfinder (http://eatworms.swmed.edu/~tim/primerfinder), primers flanking the $(CA)_n$ repeats were designed from sequence data from the identified clones (fig. 2.6a). The sequence containing the primer, flanking sequence and entire repeat (400-500bp in length) (fig. 2.6b) was BLAST-searched against the complete, incomplete and the nonredundant nucleotide databases (http://www.ncbi.nlm.nih.gov/BLAST) to confirm the degree of homology to other sequences in the database. The set of sequences from the BAC or cosmid clones were



(a)

CTCCCAAAGTGTTGGGATTGCAGGTGTGAGCCACTGCACCTGGCCCCAAAGATGCT
TTTTGATATATGTCACACACACACACACACACACACACGCCCAGCAAAAGATGTGCA
AGGATAGTATAGTAATGAAACTCGGGGTGCACATGAAATGTGAGGTCTGTATCACT
CAATATTTTTGTTGTTTGAGATGGAGTCTCTCTCTGTCACCCAGGCTGGAGTGCAGT
GGCATGATCTCAGTTCGCTGCATCCTTTGCCTCCCAGGTTCAGGTGATTCTCCTGCCT
CAGCCTCCTGAGTAGCT

Primer design (Primerfinder)
PCR-amplify repeat motifs

(b)

Forward primer
CCCAAAGTGTTGGGATTGCAGGTGTGAGCCACTGCACCTGGCCCCAAAGATGCT
TTTTGATATATGTCACACACACACACACACACACACACGCCCAGCAAAAGATGTGCA
AGGATAGTATAGTAATGAAACTCGGGGTGCACATGAAATGTGAGGTCTGTATCACT
CAATATTTTTGTTGTTTGAGATGGAGTCTCTCTCTGTCACCCAGGCTGGAGTGCA

5'-TGCACTCCAGCCTGGGTGAC-3'
Reverse complement
reverse primer

PCR-amplify in panel of unrelated individuals

Figure 2.6 Diagrammatic representation illustrating the generation of $(CA)_n$ markers from recombinant cosmids and BACs.
(a) Sequence containing a $(CA)_n$ repeat from the PFHBI locus, cloned into a BAC/cosmid (b) $(CA)_n$ repeat containing sequence flanked by the forward and the reverse primers.

only selected for marker development if it had a 100% homology to the clone from which it originated.

The selected 400-500bp sequence was copied from the clone sequence and using the "paste" option the sequence was placed into the PRIMERFINDER window. Using the "next" option, a new window was activated which allowed the parameters for the primer design to be set, namely, primer length, GC content, Ta, and $MgCl_2$ concentration. Using the same window, the forward or the reverse complement primer of the marker could be generated. Once the appropriate button was selected the "submit" option was activated. A separate web page containing all the possible primers, their sequence, Ta, and hairpin looping and self complimentarity was given. The same procedure was performed for the reverse primer (reverse complement). In addition, another two windows could be activated, one in which the forward primer and a second window in which the reverse was pasted. Using the "submit" option, the complimentarity between the two primers was determined.

2.10.3.2 Chromosome 19 somatic cell hybrids and DNA isolation from recombinant clones

Chromosome 19 somatic cell hybrids

The two somatic cell hybrids, ORIM7-1 and 1219G2, as described in section 1.4.2.1, were obtained from the Dept of Human Genetics, University of Amsterdam, The Netherlands (Schonk *et al.*, 1989). Somatic cell hybrid ORIM-7 consisted of a human t(X;19) (Xqter- Xp21;19q13.1-qter) chromosome and a few additional human chromosomes in a hamster Wg3-h background. Hybrid 1219G2 was obtained from the

fusion of hamster A3 cells and parental lymphoblastoid cells carrying a reciprocal t(12;19) translocation. Hybridisation and cytogenetic studies revealed that that 1219G2 carries human 19q13.1-qter::12pter-q15 segment (Schonk *et al.*, 1989).

The somatic hybrids cell lines were maintained in Dulbeco's modified Eagle's medium supplemented with 15% fetal calf serum, 100µg/ml penicillin G, and 100µg/ml streptomycin. The hypoxanthine-guanine phosphoribosyl transferase (HPRT)-positive or the HPRT-negative somatic cell hybrids were selected in DMEM containing either 13.6µg/ml hypoxanthine, 0.22µg/ml aminopterine and 9.7µg/ml thymidine (HAT) or 12µg/ml thioguanine, respectively (Schonk *et al.*, 1989). The DNA was extracted as described by Corfield *et al.*, (1993)

BAC and cosmid DNA isolation (Mini prep)

BAC and cosmid clones were cultured and plated as described in section 2.9.1.1 (Appendix I). In the case of BAC clones, the antibiotic chloramphenicol (50mg/ml) was used and in the case of the cosmid clones, the antibiotic kanamycin (Kn) (50mg/ml) was used. Single colonies were selected and grown overnight in 10ml LB with the appropriate antibiotic. The DNA was extracted as follows: Each tube was spun at 3500rpm at room temperature in a Sorvall centrifuge (Du Pont Instruments, model RC-5B) for 10 min (Birnboim, 1983). The resulting pellet containing bacterial cells was gently resuspended in 200µl of chilled solution I (GTE) (Appendix I). The next step was the addition of 400µl of freshly prepared solution II (0.2M NaOH; 1% SDS) (Appendix I) to the suspension. To this 300µl of solution III (3M KAc, 11.5% CH_3COOH) (Appendix I)

were added and the tube was centrifuged for 6 min at 13000rpm in a benchtop centrifuge (model Microfuge^R Lite, Beckmann Instruments, California, USA) at room temperature. Four hundred microlitres of supernatant were mixed with 200µl phenol and 200µl of chloroform-isoamyl alcohol, the mixture was centrifuged at 13000rpm for 5 min in a benchtop centrifuge (model Microfuge^R Lite, Beckmann Instruments, California, USA) and the aqueous top phase was removed. The phenol/isoamyl chloroform step was repeated until the aqueous phase was clear. The final aqueous phase was mixed with 800µl of 100% ethanol and the mixture was incubated at –20°C for 1 hr. The tubes were centrifuged at 13000rpm for 10 min in a benchtop centrifuge (model Microfuge^R Lite, Beckmann Instruments, California, USA) and the DNA pellets were washed with 70% ethanol and air-dried. Forty five microlitres of 1x TE and 5µl (10mg/ml) of RNAse solution (Sigma Chemical Company, St Louis, Missouri, USA) were added to the dried pellet and this was incubated at 37°C for 1 hr. The DNA was stored at 4°C until required.

2.10.3.3 PCR optimisation

The primers from the novel $(CA)_n$ STRs generated from the BAC clone sequence (fig. 2.6) were used in the PCR amplifications using the conditions shown in table 2.2 in section 2.6. Each PCR reaction was optimised with respect to the varying conditions and parameters such as Ta, extension time and $MgCl_2$ concentration (table 2.2). In addition, a set of cosmid and BAC clones were PCR amplified using different primer sets in order to verify that the primer set matched the purchased DNA clone from Research Genetics, USA. The following five control reactions were set up: PCR amplifications were prepared using each set of new primers with the isolated DNA from recombinant BAC or

cosmid DNA clones from which the primers were generated; another control was set up, which used DNA isolated from recombinant cosmids or BACs from other DNA clones at the PFHBI locus. To verify that the primers came from a recombinant DNA clone from the chromosome 19q region, the chromosome 19q radiation hybrids were amplified. Genomic DNA was amplified with the novel primer sets as a positive control and as a negative control the DNA was replaced by water. The PCR protocol detailed in section 2.6.1 was used and 2µl of the PCR products mixed with 1µl bromophenol loading buffer were electrophoresed on a vertical (200mm x 150mm x 0.4mm) 12% non-denaturing polyacrylamide gel (Appendix I), with a sample of 1µl λ *Pst*1 used as a molecular size marker (Appendix I). The gel was silver stained as described in section 2.7.2 (Appendix I).

2.10.3.4 Determination of polymorphic nature of marker

The polymorphic nature of the $(CA)_n$ repeat markers were determined in the same manner as the $(A_3G)_n$ repeat marker (section 2.9.6).

2.10.3.5 Genotyping

Once the PCR conditions were optimised, as judged by the generation of the expected size product seen after silver staining, a radioactive PCR was performed using the conditions described in section 2.6.2. A volume of 3-5µl of the PCR products were electrophoresed on a 6% denaturing polyacrylamide gel (390mm x 290mm x 1mm) as described in section 2.7.3 and Appendix I. After this, genotyping of the novel markers

was performed in the PFHBI families, in an attempt to refine the existing recombination breakpoints.

2.10.4 Submission of novel genetic markers to STS database

Novel $(CA)_n$ repeat genetic markers developed during this study were submitted to the STS (sequence tag site) database (http://www.ncbi.nih.nlm.gov/STS). There are two sets of file types for STS submissions; one is for the sequence data and one for any mapping data. The former was used for the submissions for this study (batch-sub@ncbi.nlm.nih.gov). After submission, there are validation checks performed by dbSTS curators. If the validation checks are successful, the dbSTS curator sends a list of dbST IDs and Genbank accession numbers to the submitter.

2.11 DATABASE SEARCHES FOR GENE TRANSCRIPTS

2.11.1 Transcript Searches

At the start of this study, very few genes at the PFHBI locus had been annotated and deposited into the publicly available databases but a number of ESTs had already been placed by other researchers (fig. 2.7). In the search for the PFHBI gene, it was important to find as many genes as possible in the region, in order to prioritise them as potential candidate genes.

2.11.1.1 Retrieval of ESTs

During the course of the study, the dbEST (http://www.ncbi.nlm.nih.gov/dbEST) was established and the data made publicly available. In addition, ESTs from the EST database for each chromosome were also being assembled and deposited into the Genemap99 database (http://www.ncbi.nlm.nih.gov/Genemap99). Concurrently, the group at LLNL, Livermore, USA involved in sequencing chromosome 19, was depositing ESTs at their site (http://www.bio-llnl.gov/genome). As part of the project, all the ESTs mapping to the PFHBI critical region were retrieved from the available public databases and integrated into one map (fig 2.7a).

2.11.1.2 Clustering of ESTs

Since these retrieved ESTs were typically only 200-300 base pairs in length, a BLAST-search of their translations against the protein database might not result in significant protein matches. Therefore, to obtain longer sequences, these ESTs were extended using STACKpack from the STACK database (http://www.sanbi.ac.za/STACK) and UNIGENE (http://www.ncbi.nlm.nih.gov/UNIGENE). The procedure for the STACK analysis is as follows: the STACK web page has one sequence input window. The accession number of the EST from a database was retrieved, using the "copy" option, and inserted, using the "paste" option, into the "subject" window in the STACK web page. The information was then submitted and the STACK results containing all the cluster homologies with their scoring values were returned in a separate web page.

The procedure for the UNIGENE database was as follows: The accession number of an EST from a database was retrieved, using the "copy" option, and inserted, using the "paste" option, into the "subject" window in the UNIGENE sequence web page. The information was then submitted and the output was returned in a separate web page which contained protein homology, tissue expression profiles, homologene, all the ESTs with their orientation and also access to a number of different databases.

The clustering programme generated a number of STACK clusters and the STACK cluster with the highest scoring cluster value was selected for the BLAST search against the SWISS-PROT database (fig. 2.7b); whereas a search of the UNIGENE database using the retrieved ESTs produced direct protein homologies to proteins in the SWISS-PROT database (fig. 2.7e).

2.11.1.3 Positioning of ESTs on contig NT_011109

In silico (computer-based) mapping was performed via the Internet (US/University of Western Cape (UWC) internet service provider), which involved finding the position of each EST on the chromosome 19 contig NT_011109 using the BLAST search tool. A BLAST search is performed when the electronic version of a sequence is pasted into the window of the database of interest and the "submit" button is activated. For example, in this case an EST, the sequence is pasted into the window of the Human Genome database, after which a BLAST search against this database was performed in order to verify the bp positions of the EST on chromosome 19 contig NT_011109 within the PFHBI locus (http://www.ncbi.nlm.nih.gov/human genome) (fig. 2.7d).

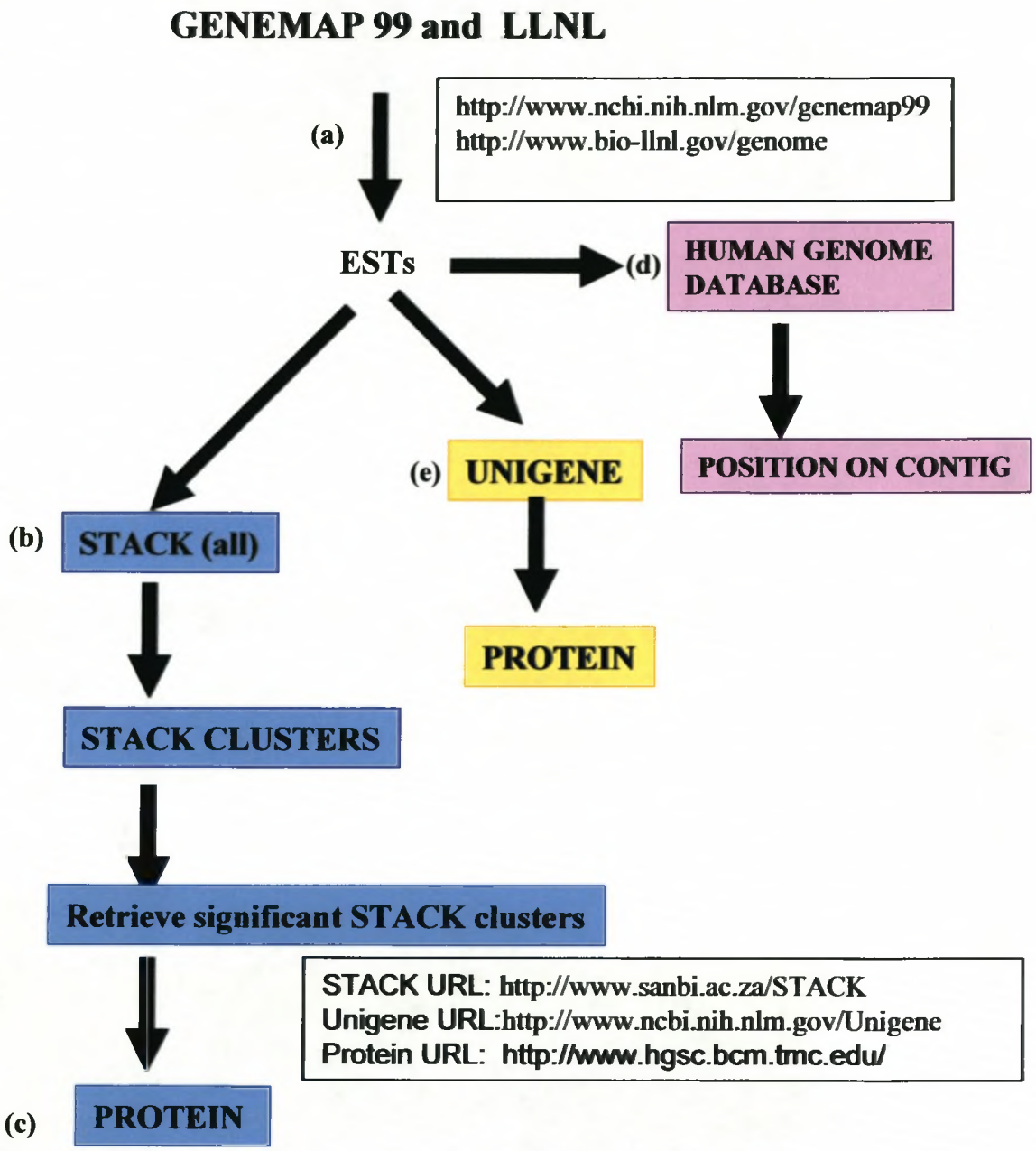


Figure 2.7 Illustration of the steps involved in the search for gene transcripts with the view to finding novel genes. (a) The retrieval of ESTs from the EST databases (b) A BLAST search of the retrieved ESTs against the STACK database (c) the STACK cluster matches against the protein databases. (d) A BLAST search of against the Human Genome database (e) A BLAST search of the ESTs against the UNIGENE database.

2.11.2 Protein homology searches

The retrieved STACK clusters were BLAST-searched against the SWISS-PROT protein database using BCM Searchlauncher (http://www.hgsc.bcm.tmc.edu/) for possible protein homologies. Both STACK and UNIGENE databases contained tissue expression profiles for the different ESTs and this information was used to detect the tissue profile expression for each EST and gene placed at the PFHBI locus.

2.11.3 Interpretation of BLAST data

Expectation values (E values) are the scores reported by a similarity programme such as BLAST on their output file. The E-values should be treated like P values from traditional types of statistical tests and represent the number of times that a match with observed similarity score would occur by chance if the database actually contained no sequences that were truly similar to the query sequence. Generally, E values of zero or less than 10^{-100} are assumed to be significant matches (Brown, 2000) and E values in this range was used in this study.

2.12 INTEGRATING OF PHYSICAL MAP

To facilitate the search for the PFHBI-causative gene, a high resolution, integrated physical of the PFHBI target region was required. Two approaches were used in this study: the first was a bioinformatic approach, which involved the retrieval of all available information from publicly available databases and then applying bioinformatic tools to

the retrieved data. The second approach was the use of laboratory based experimental techniques aimed at physical mapping of loci (fig. 2.8).

2.12.1 Data-mining

The sequence data and Genbank accession number of the BAC and cosmid clones spanning the 4Mb target region were retrieved from the LLNL database and the JGI site (http//:www-bio.llnl.gov/genomes/; http://www.jgi.doe). This step was followed by the retrieval of ESTs from the EST databases, as described in section 2.10.

Information about the STSs, published/database genetic markers and the annotated genes were retrieved from the NCBI database (http://www.ncbi.nlm.nih.gov/nucleotide). The positions of the BAC clones, cosmid clones, genetic markers and genes on the contig NT_011109 were found by performing BLAST searches (Altschul *et al.*, 1990) against the human genome database using the Genbank accession numbers or the FASTA formatted sequences of the clones. ESTs corresponding to the respective genes and the expression profiles of each gene were obtained by searching the UNIGENE database (fig. 2.8).

Since a number of annotated genes were being deposited into the database, BLAST searches of the Human Genome Database were performed to find the position of these genes on the PFHBI contig. In order to find a gene contained in a BAC clone, the Genbank accession number or the FASTA formatted sequence of the gene, was submitted

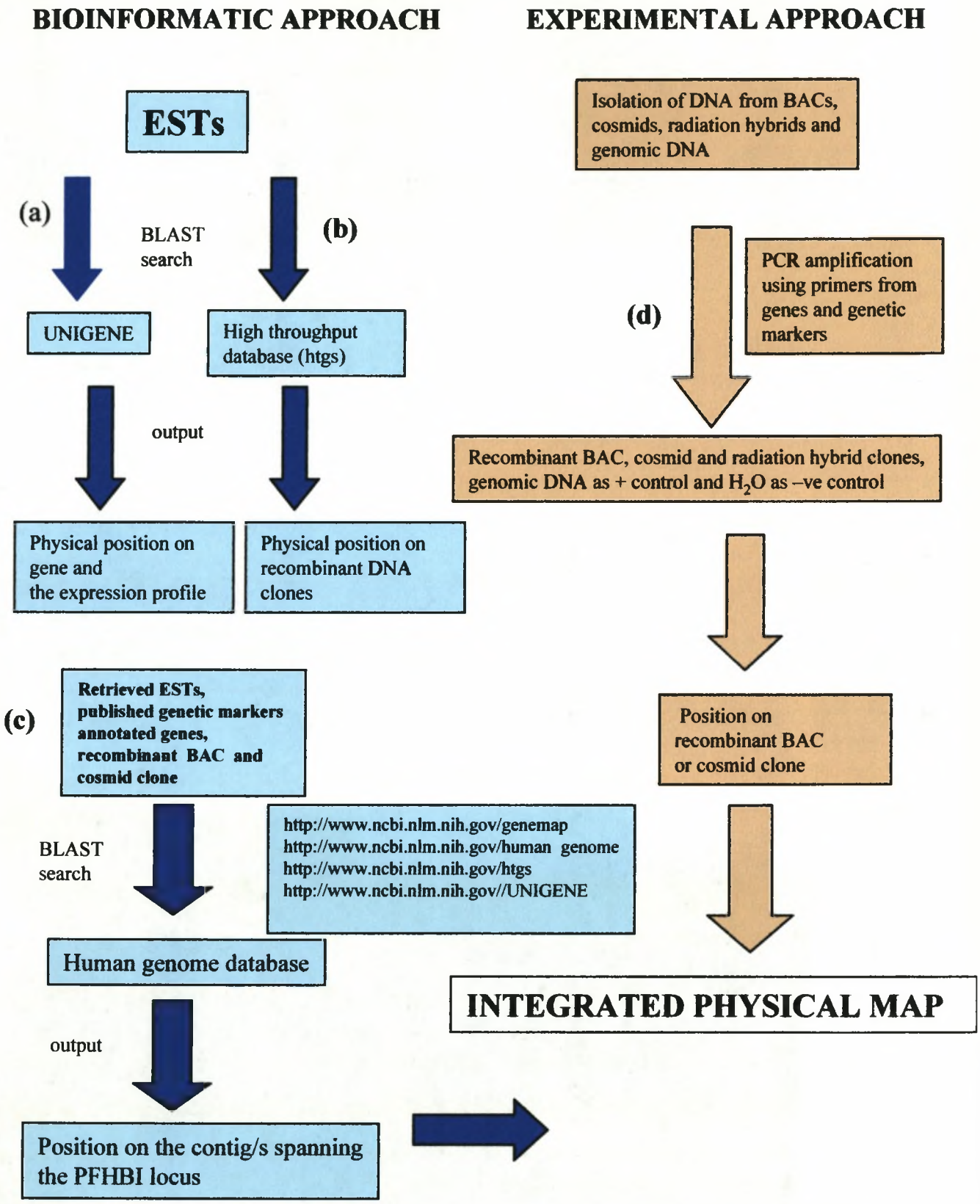


Figure 2.8 Schematic diagram illustrating the two approaches used to produce an integrated physical map.

(a) Linking of the ESTs to genes and the determination of the expression profiles each gene, (b) Determination of the physical position of the ESTs on the recombinant BAC/cosmid clones. (c) Placement of the sequences of ESTs, published genetic markers, annotated genes and recombinant BACs and cosmids clones on the PFHB contig/s (d) PCR amplification of published primers with recombinant BACs, cosmids and radiation hybrids clones.

to the high throughput database (htgs) (http://www.ncbi.nlm.nih.gov/htgs). Additional information regarding annotated genes was also obtained from ENTREZ at NCBI (http://www.ncbi.nih.nlm.gov), Ensembl, (http://www.ensembl.org), Genbank (http://www.gdb.org) and LOCUSLINK (http://ncbi.nih.nlm.gov/LOCUSLINK).

2.12.2 Experimental approach

PCR amplification was used to verify the position of the genetic markers on the BAC and cosmid clones, chromosome 19 radiation hybrids and genomic DNA. The selected DNA clones spanning the target region were firstly extracted using the protocol outlined in section 2.9.1.1 (Appendix I). Thereafter, amplification using the genetic marker primer sets from tables 2.1 and 2.2 on the various DNA sources were used as described in section 2.6.1. The PCR products were electrophoresed on a 1-2% agarose gel as described in section 2.7.1. The data obtained from the data-mining and the experimental approaches (sections 2.12.1 and 2.12.2) were collated and used to construct the integrated physical map. The subsequent step involved the analysis and prioritisation of the available annotated and predicted genes, present at the PFHBI locus, for mutation screening.

2.13 ANALYSIS OF PREDICTED AND ANNOTATED GENES

2.13.1 Criteria for prioritisation of PFHBI candidates

All the genes mapped to the PFHBI locus are candidates by virtue of their position. However, some genes are functionally more attractive candidates because of additional

features, such as expression in cardiac tissue, association with cardiac function and cardiac physiology, implication in cardiac disease, involvement with development, involvement in the apoptotic pathway, and presence of triplet repeat motifs present either in the exonic, intronic or promoter regions.

2.13.2 Position, function and expression profile

Information regarding the position, function and expression profile of both annotated and predicted genes present at the PFHBI locus was retrieved from the various databases at NCBI, such as UNIGENE (http://www.ncbi.nlm.nih.gov/UNIGENE) and Locuslink (http://www.ncbi.nlm.nih.gov/Locuslink) was obtained.

Known genes are defined as transcripts identified by experimental analysis, which are expressed *in vivo* (Stein, 2001). In addition, computer predicted genes for which there is experimental data to support that it has an *in vivo* biological function are also catalogued in HGP as a gene (Burge and Karlin, 1997).

Predicted genes which had been identified using bioinformatics similarity searches, namely, BLASTN (http://www.ncbi.nlm.nih/BLAST) and Genomescan (http://www.genomescan.gov) were retrieved as part of this study and prioritized, based on their putative functions.

2.14 GENE SEARCHING AND MUTATION SCREENING

A number of genes were prioritised based on the criteria described in section 2.13.1, as excellent candidate genes for PFHBI. In this study, using specific and significant criteria from section 2.13.1, five genes namely, *KCNA7*, *KIR2.4*, *BAX, LIN-7B and GSY1* were selected for screening for disease-causing mutations in PFHBI.

The presence of a CTG repeat expansion in the PFHBI families was investigated by Schalling of the Karolinska Institute, Sweden using the repeat expansion detection (RED) method (Schalling *et al.*, 1993) (section 2.14.1).

Professor C.T. Basson from the Department of Medicine (Cornell University Medical School, New York, USA) contacted the PFHBI group because he had isolated the cDNA of a G protein, which he had mapped to chromosome 19q13.3. In a collaborative agreement, the PFHBI group used the Basson's clone to probe the DNA clones spanning the PFHBI locus for the presence of the gene (section 2.14.2).

Dr P. Coucke from the University of Antwerp, Belgium contacted the PFHBI group because he suggested an investigation of the PFHBI locus for a *Cx*, since *Cxs* were thought to be strong candidate genes for PFHBI (section 1.2.2.3). PCR amplification using a set of degenerate primers (fig. 2.2), designed and synthesised by Coucke's group, DNA from clones spanning the PFHBI locus and a selected panel of PFHBI individuals (see fig. 2.1) to search for the presence of a *Cx* (section 2.14.3).

In the case of *KCNA7*, the PFHBI group contacted Professor G. Chandy from the Physiology and Biochemistry Department (University of California, Irvine, USA) because his group had localised, characterised and sequenced *Kcna7* of *M. musculus* (Kalman *et al.*, 1998) and, in addition, had placed the human orthologue, *KCNA7*, at the PFHBI locus. A collaboration with this group resulted in the characterisation of the human *KCNA7* (Bardien-Kruger *et al.*, 2002) (section 2.14.4).

In the case of *HRC*, Professor Brian Black from the Cardiovascular Institute (University of California, San Francisco, USA) contacted the PFHBI group in 1999 and offered to sequence the HRC gene, including the 5' and 3' UTRs, as well as the complete intronic sequences of one affected and one unaffected first degree relative of PFHBI individuals from each of the three pedigrees.

2.14.1 Search for CTG repeat expansion

The RED technique can identify potentially pathogenic repeat expansions without prior knowledge of chromosomal location. Genomic DNA is used as a template for a two step cycling process that generates oligonucleotide multimers when expanded trinucleotide sequences are present, as in the case of DM and Huntington's disease (fig. 2.9) (Schalling *et al.*, 1993). Five DNA samples from affected PFHBI individuals were analysed by the Schalling group for possible CTG repeat expansions.

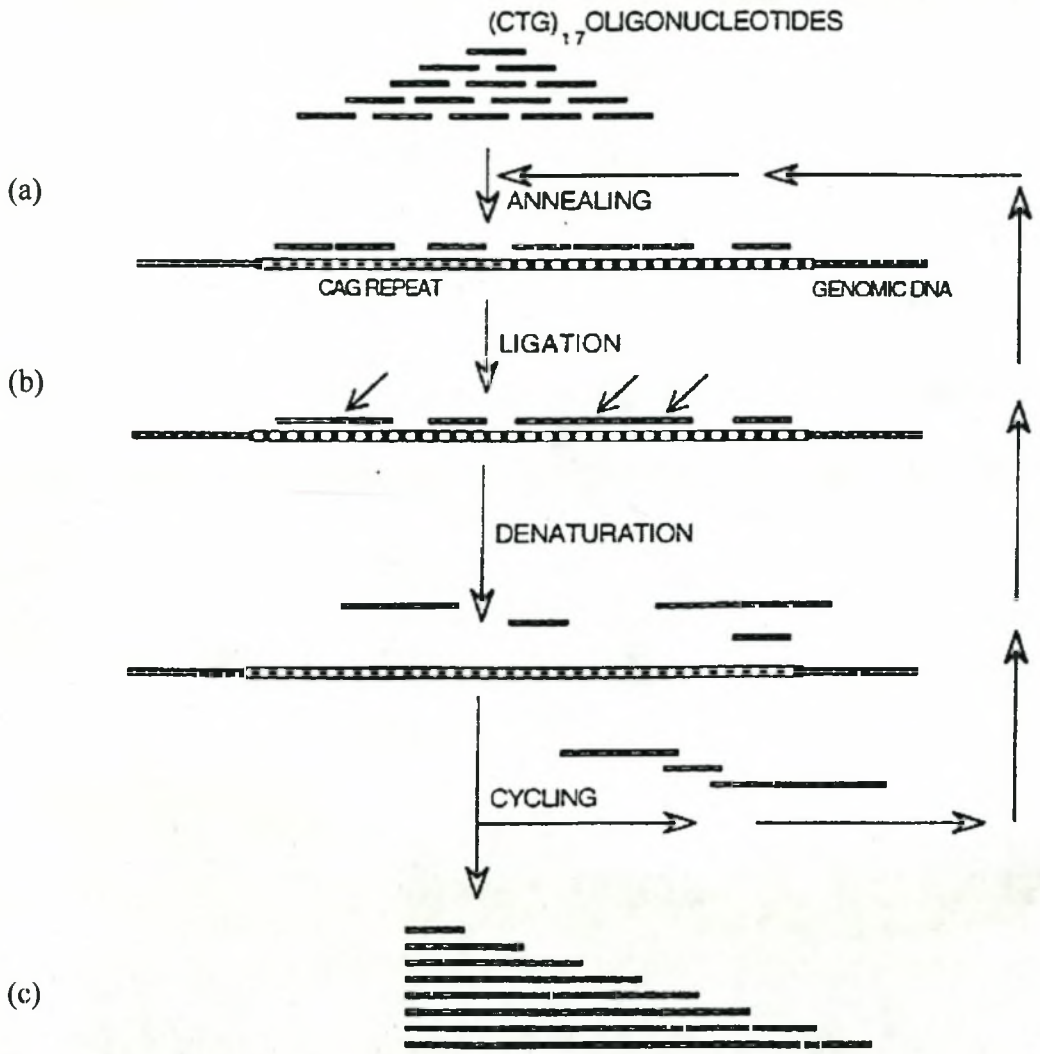


Figure 2.9 Principle of repeat expansion detection (RED) technique.
The diagram illustrates the different steps of the RED technique which detects expanded trinucleotide $(CAG)_n$ repeats in genomic DNA. (a) Single-stranded genomic DNA containing a CAG repeat and a $(CTG)_{17}$ oligonucleotide is hybridised at a temperature near its melting point. (b) Thermostable ligase binds covalently adjacent oligonucleotides, producing multimers of the oligonucleotide $(CTG)_{17}$. The ligation step is then repeated several hundred times. (c) A mixed population of single-stranded multimers are produced. The detection of single-stranded multimers is achieved by size separation on a denaturing polyacrylamide gel, which is subsequently electrotransferred, hybridised and then probed with a ^{32}P-labeled complementary oligonucleotide (Schalling *et al.*, 1993).

2.14.2 G protein-encoding gene

A human guanine nucleotide exchange factor *p115-RhoGEF* cDNA clone (Genbank Acc No. U64105) was obtained from C.T. Basson and was used to probe a dot blot of cosmids, BACs, YACs, DNA clones and chromosome 19 radiation hybrids (section 2.10.1.3) spanning the PFHBI locus for the presence of G protein gene. The p115-GEFcDNA insert, 800 to 900bp in size, cloned in vector pCR2.1 (Stratagene, USA) (fig. 2.10), was labeled (Appendix I). This was followed by hybridisation with the labeled probe and subsequent washing (Appendix I). The hybridised dot blot was exposed to an X-ray film (Cronex-4, Protea Medical) overnight at room temperature. The X-ray film was developed as described in 2.7.4.

2.14.3 *Cx*

PCR using degenerate Cx primers

The presence of a *Cx* at the PFHBI locus was investigated by the amplification of all the cosmids, BACs and the chromosome19q radiation hybrids with the degenerate primers, using the protocol described in section 2.6.1 (fig. 2.2). The set of Cx degenerate primers, were obtained from P. Coucke, University of Antwerp, Belgium, which were designed by the alignment of six *Cxs*, namely, *Cx26* (M8849), *Cx37* (M96789), *Cx40* (U03486), *Cx43* (X52947), *Cx45* (U03493), and *Cx50* (U34802), using the CLUSTALW alignment programme (fig. 2.2) (personal communication, Coucke).

The PCR products were electrophoresed on a 1% agarose gel, the expected 200bp product was cut out of the gel as described in section 2.7.1 and extracted using a GFX PCR DNA extraction kit (Amersham, UK) according to the manufacturer's protocol (section 2.6.3.1). Automated sequencing of all the PCR products was performed at the DNA Sequencing Core Facility, University of Stellenbosch (US).

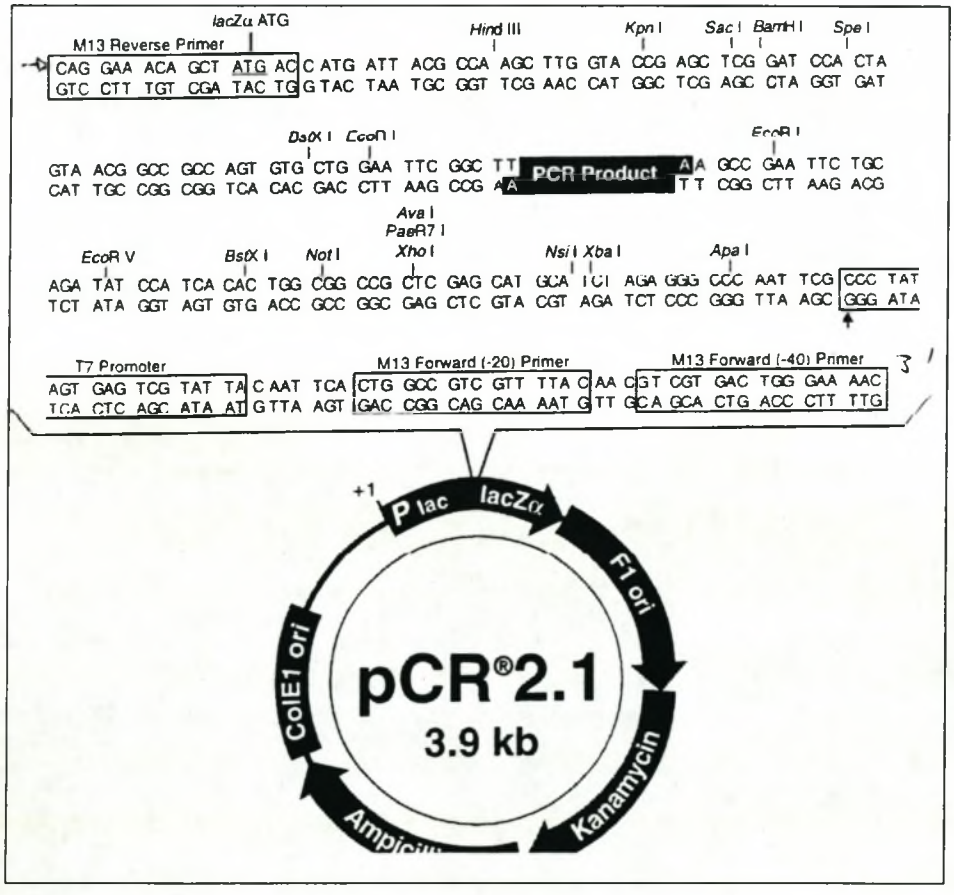


Figure 2.10 Map of vector pCR2.1 (Stratagene, USA) **containing the *p115-GEF* cDNA insert.**

Analysis of sequence of PCR products

A protein homology search was done against the non-redundant protein database (http://www.ncbi.nih.nlm/BLAST/nr) using the Cx-amplified sequences obtained, to identify significant protein matches.

A BLAST2 (http://www.ncbi.nih.nlm/BLAST2/) homology search was performed which involved the comparison of the *Cx*-amplified sequences with BAC or cosmid sequence spanning the PFHBI locus. The electronic versions of the genomic clone-insert sequences and Cx-amplified sequences were aligned to each other using the "Blast 2 sequences" option of the *BLAST* algorithm. The procedure for the alignment analysis is as follows: the BLAST2 sequence web page has two sequence input windows. The accession number of a genomic clone-insert was retrieved, using the "copy" option, from a database and inserted, using the "paste" option, into the "subject" window in the BLAST2 sequence web page. The sequence of the *Cx*-PCR product was inserted into the "query" window using the "copy and paste" options described previously. The information was then submitted and the alignment results were returned in a separate web page.

2.14.4 *KCNA7*

The characterisation of the coding sequences of genomic DNA sequences

The sequence of the human *KCNA7* was obtained by performing a BLAST search using the mouse *Kcna7* cDNA (AFO32099) against the Genoscope database (http//www.genoscope.org) and in addition a search against Genbank was carried out with the inferred human *KCNA7* sequence (Bardien-Kruger *et al.*, 2002). The coding

sequence of *KCNA7* was deduced by aligning the mouse cDNA sequence (AFO32099) with the programme EST2GENOME (http://ftp.sanger.ac.uk/pub/EMBOSS).

Mutation screening

As art of this study, two sets of overlapping of primers were designed from *KCNA7* sequence for exon 1 and five sets for the larger exon 2 (table 2.3) (fig. 2.11). Using the primer sets for exon 1 and exon 2, the coding and flanking intronic regions of affected individuals from pedigrees 1 (1.V.10), 2 (6.IV.12 and 9.IV.5) and 5 (5.III.1), unaffected individuals from pedigrees 1 (1.V.6) and 2 (9.III.6) and a married-in (13.II.8) were screened for pathogenic mutations using PCR-SSCP analysis and direct sequencing as described in sections 2.7.2 and 2.6.3, respectively.

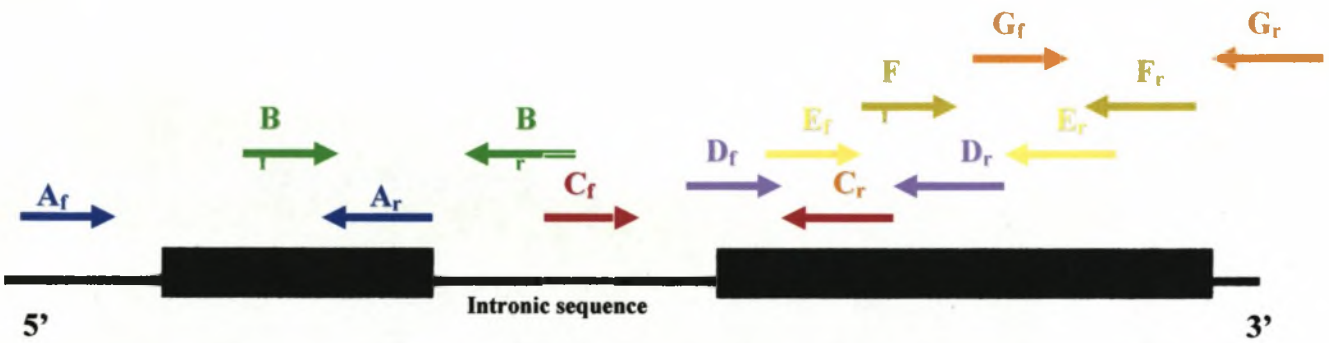


Figure 2.11 Schematic representation of the overlapping primers designed for the screening of *KCNA7*.
The filled black boxes correspond to the exonic sequences, and the black lines correspond to the intronic sequences. The overlapping primer sets A and B designed for the amplification of exon1 and its flanking intronic sequences are shown. The overlapping primer sets C, D, E, F and G designed for the amplification of exon 2 and its flanking intronic sequences are shown (Bardien-Kruger *et al.*, 2002).

The design of *KCNA7* primers, the mutation screening of the exonic sequences of the KCNA7 and the investigation of the two SNPs in exon 2 of *KCNA7* were investigated as part of this study.

2.14.5 *BAX, KIR2.4, LIN-7B* and *GSY1*

The published or web based primer sequences of *BAX, KIR2.4* (fig. 2.12*), LIN-7B* and *GSY1*, as shown in table 2.2, were used to amplify the exonic and flanking sequence of genes, analysed using the PCR-SSCP. Automated sequencing of all the PCR products was performed at the DNA Sequencing Core Facility, US and the sequence analysis and the CLUSTALW alignment were performed as described in sections 2.8.1 and 2.8.2.

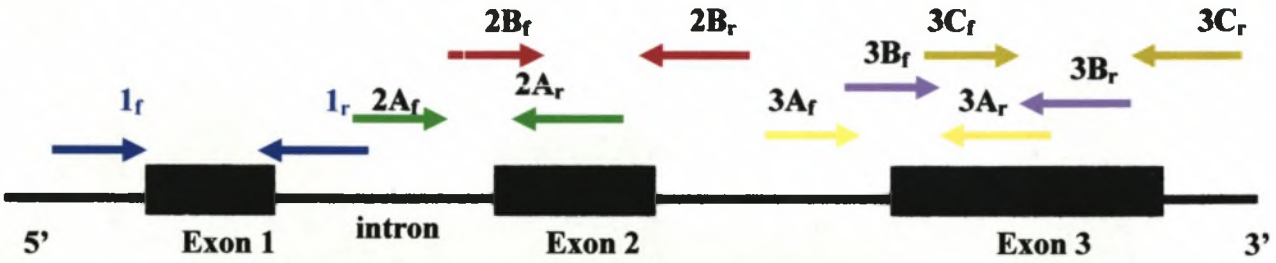


Figure 2.12 Schematic presentation of the positions of overlapping primers for the screening of *KIR2.4.*
The shaded black boxes correspond to the exonic sequences, and the black lines correspond to the intronic sequences. The two overlapping primer sets **2A** and **2 B** were designed to screen exon 2 and its flanking intronic sequences. The three overlapping primer sets **3A, 3B and 3C** were designed to screen exon 3 and its flanking intronic sequences.

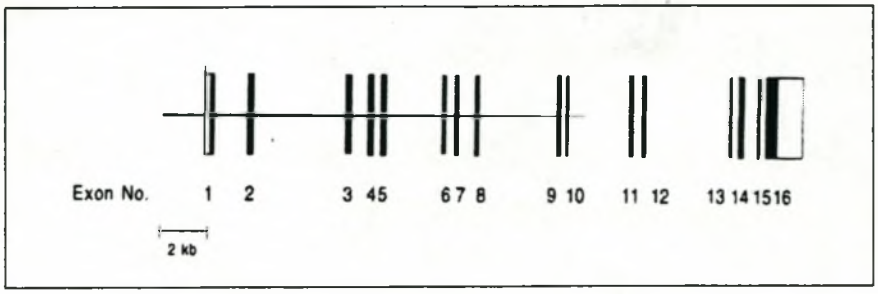


Figure 2.13 Schematic diagram of the structure of the *GSYI* gene.
Each exon is represented by a box and is numbered. The filled boxes correspond to translated sequences, and the open segments of a box correspond to untranslated sequences **(Ohro *et al.*, 1995)**.

*A*s part of this study the 5'UTR, the exonic and the flanking regions of exons 1-3, 6-10, 13-16 of *GSYI* (fig. 2.12) were screened for mutations by direct sequencing of the PCR products obtained using the PFHBI panel (fig. 2.1).

CHAPTER 3

RESULTS

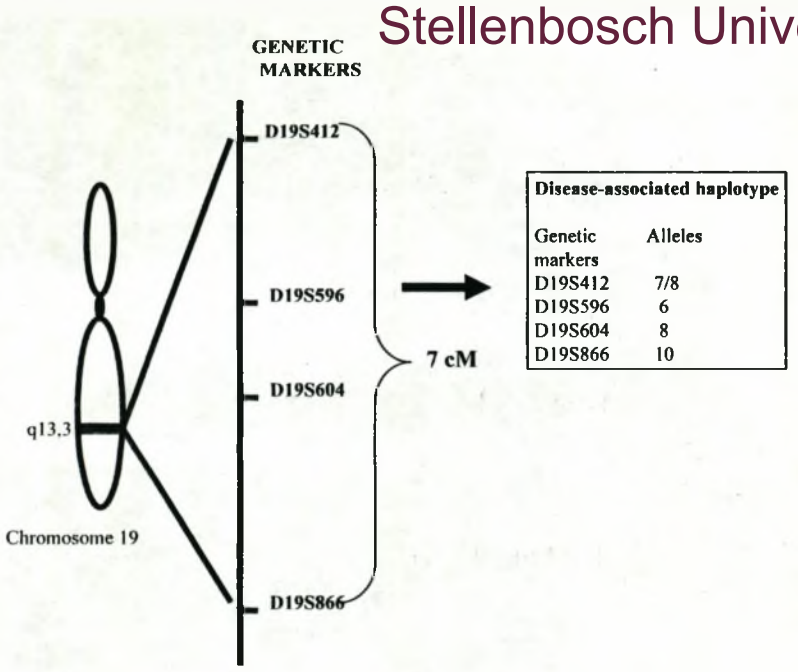
INDEX

3.1 REDUCTION OF THE PFHBI LOCUS BY GENETIC FINE MAPPING USING POLYMORPHIC MARKERS

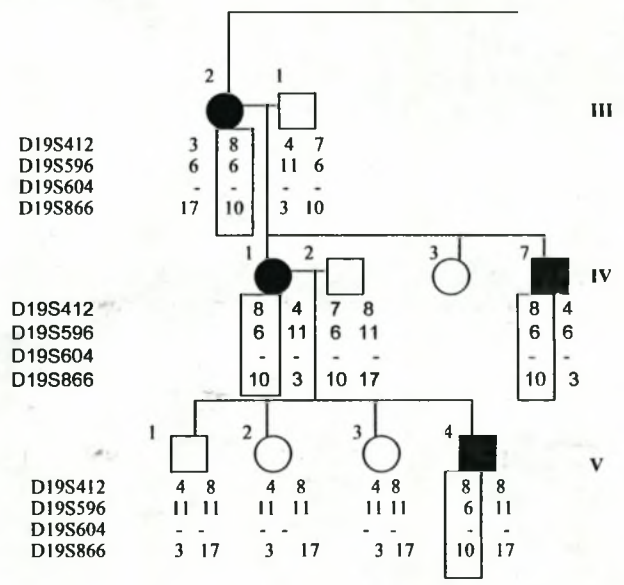
3.1.1 PFHBI disease-associated haplotype

Previous studies conducted by other researchers fine mapped the PFHBI locus to a 7cM region on chromosome 19q13.3 between genetic markers *D19S412*, at the centromeric limit, and *D19S866*, at the telomeric limit, thereby establishing the disease-associated haplotype in PFHBI-affected individuals in pedigrees 1, 2 and 5 (fig. 3.1) (personal communication, De Jager, 1998).

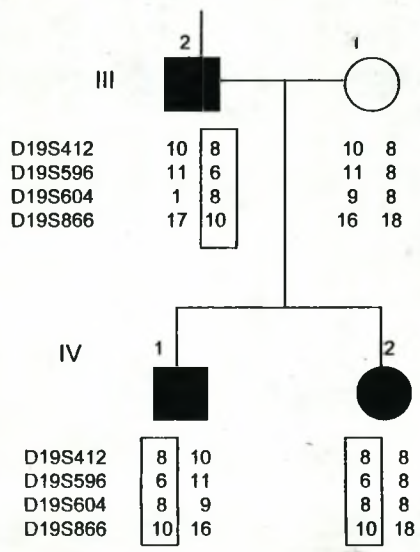
As part of a previous study, and continued as part of this study, genotyping with genetic markers *D19S412-D19S596-D19S604-D19S866* generated a disease-associated haplotype of 8-6-8-10 for the PFHBI-affected individuals shown in pedigrees 1, 2 (kindreds 7, 9 and 10) and 5 (fig. 3.1), or 8-6-(/)-10, as despite many efforts to overcome difficulties, *D19S604* was refractory to amplification in some individuals. The possibility of a mutation at the primer binding site might be the reason for the problems experienced with this particular marker in both affected and unaffected in pedigree 1 and 5 and should be investigated using the allele-specific oligonucleotide hybridisation method. The genotyping of PFHBI-affected individuals in kindred 4 of pedigree 2 with these markers produced a disease-associated haplotype of 7-6-8-10 or 7-6-(/)-10 (fig. 3.1). The inheritance of allele 7 of marker *D19S412* for PFHBI-affected individuals in kindred 4 of pedigree 2, in the disease-associated haplotype, indicated that a recombination event had occurred in the region between *D19S412* and *D19S596* in affected individuals of these



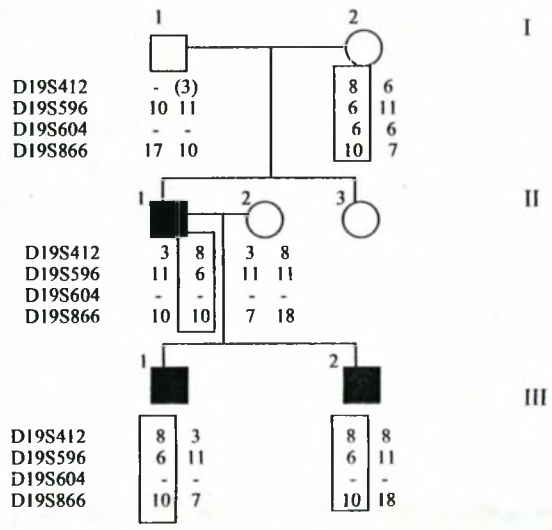
PEDIGREE 1



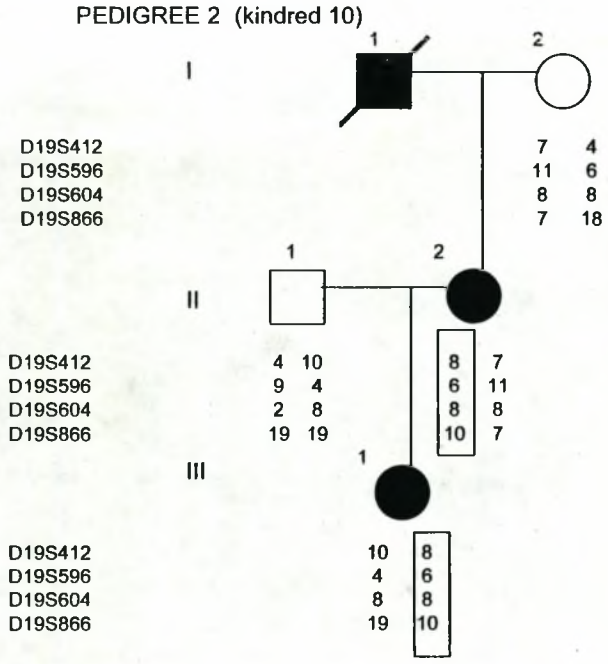
PEDIGREE 2 (kindred 7)



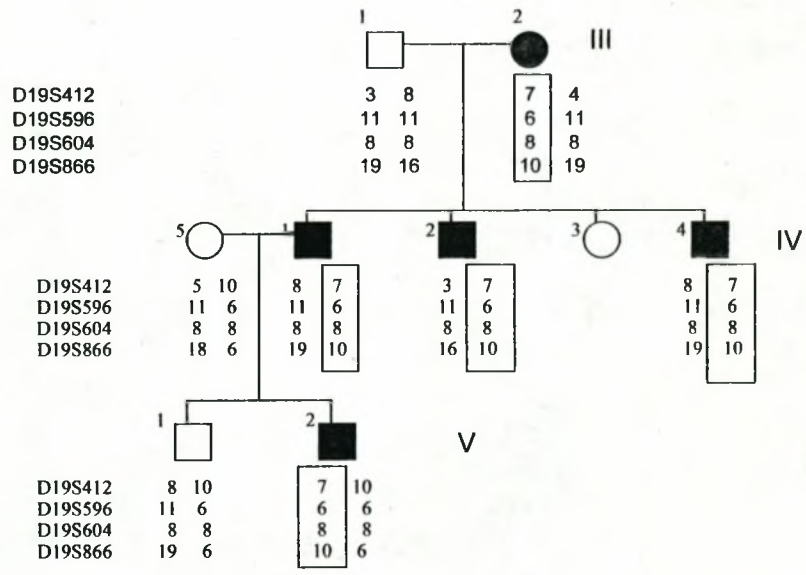
PEDIGREE 5



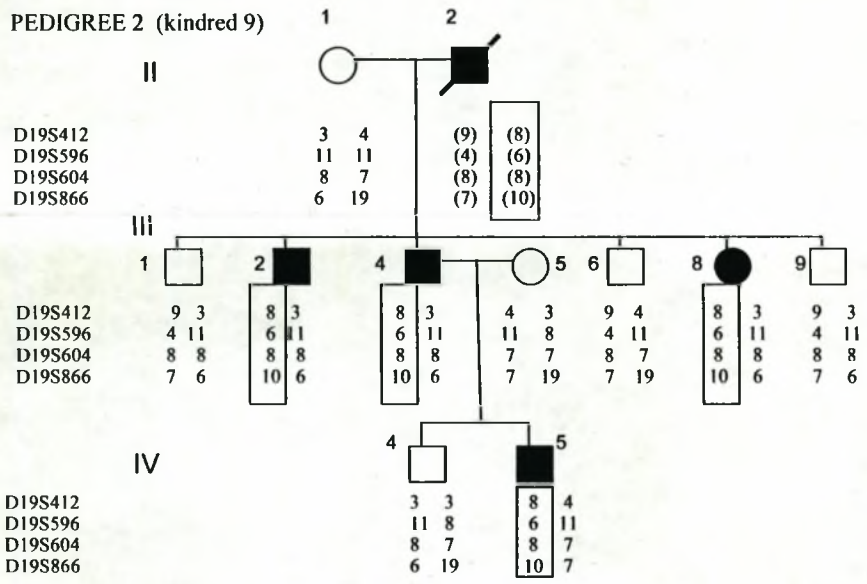
PEDIGREE 2 (kindred 10)



* PEDIGREE 2 (kindred 4)



PEDIGREE 2 (kindred 9)



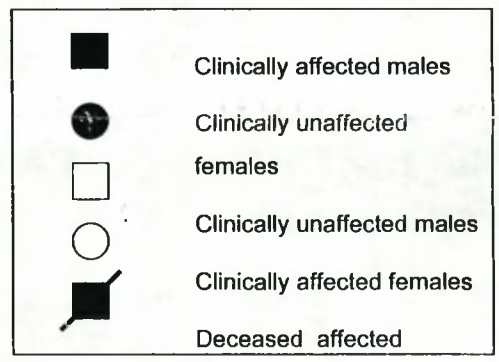


Figure 3.1 Schematic representation of the genetic map of the PFHBI locus and the disease-associated haplotype in pedigrees 1, 2 (kindreds 4, 7, 9 and 10) and 5. Pedigrees 1, 2 and 5 are shown with the disease-associated haplotype at markers *D19S412, D19S596, D19S604 and D19S866.* Alleles which have been inferred are indicated with brackets. The disease-associated haplotype in each kindred is boxed. The pedigree generations are indicated by roman numerals. *, Indicates the pedigree and the kindred in which the locus-defining recombination event has occurred.

families but not in the remaining families. Therefore, as part of this study, kindred 4 of pedigree 2 was genotyped with known and novel genetic markers, lying between *D19S412* and *D19S596*, (as described in sections 3.1.3 and 3.1.5) in order to refine the PFHBI locus. DNA samples of individuals from kindred 9 were also genotyped and used as the reference kindred.

3.1.2 Genotyping with *D19S606*

Since the database genetic marker *D19S606* was placed centromeric to *D19S596* and a recombinant event was present between *D19S412* and *D19S596* in kindred 4, as shown in fig. 3.1, *D19S606* was genotyped in the PFHBI families, as part of this study. Kindred 9 was used as the reference kindred. As presented in section 1.4.6 (fig. 1.9), *D19S606* was also the most centromeric marker of the CCD locus (de Meeus *et al.*, 1995).

From the comparison of the genotyping results with *D19S606* in kindreds 4 and 9 of pedigree 2, it was confirmed that a recombination event had occurred between markers *D19S412* and *D19S596* in kindred 4 (fig. 3.2). For example, four PFHBI-affected individuals from kindred 4 (4.III.2, 4.IV.1, 4.IV.2, 4.IV.4 and 4.V.2) carried allele 9, for marker *D19S606,* whereas the three PFHBI-affected individuals from kindred 9 (9.III.3, 9.III.4 and 9.III.8 and 9.IV.5) carried allele 6 for marker *D19S606*. This recombination event reduced the PFHBI locus to 4cM and reset the centromeric limit at *D19S606* (fig. 3.3). In addition, marker *D19S606* mapped to the same position as *D19S902* on the genetic map of chromosome 19 (fig. 3.1) (Makubalo, 2000). Genotyping of *D19S902* in

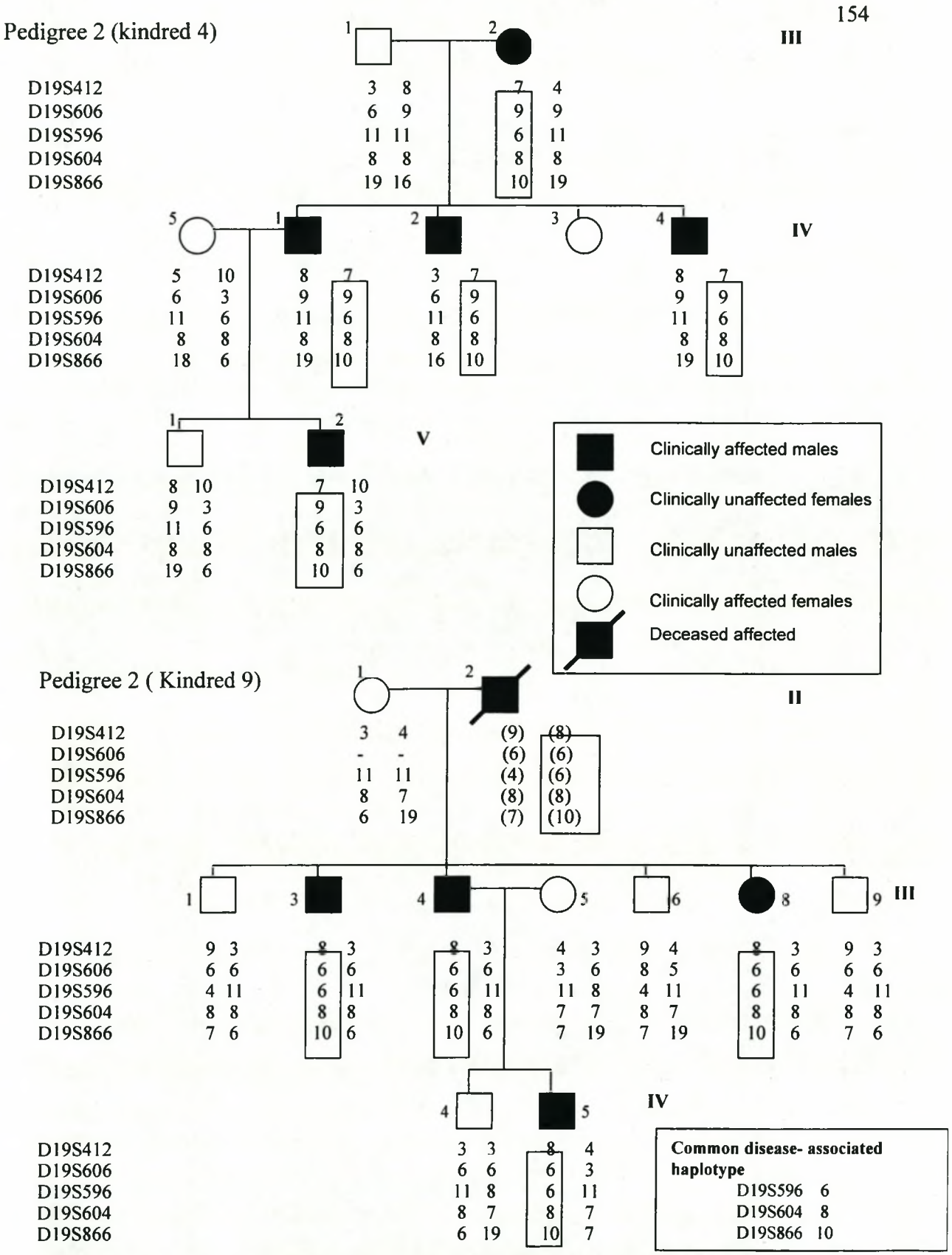


Figure 3.2 Genotyping with genetic marker *D19S606* in kindreds 4 and 9 of pedigree 2.
Kindreds 4 and 9 are shown with the disease-associated haplotype at markers *D19S412, D19S606, D19S596, D19S604 and D19S866*. Haplotypes which have been inferred are indicated with brackets. The disease-associated haplotype in each kindred is boxed. The pedigree generations are indicated by roman numerals.

kindreds 4 and 9 confirmed the recombinant event at *D19S606* (section 3.1.4, fig. 3.9).

3.1.3 Development of $(A_3G)_n$ polymorphic markers

As part of an integrated approach to reduce the size of the PFHBI locus, different members of the PFHBI group were involved in developing either polymorphic di- or tetranucleotide STR markers from the available recombinant cosmid clones spanning the PFHBI locus (fig. 3.3). As part of this study, the development of $(A_3G)_n$ repeat markers was pursued.

To this end, a Southern blot containing 11 *Sau*3A1-digested chromosome 19 cosmid clones (18618, 15217, 24493, 29395, 27268, 8946, 20381, 19161, 24596, 25235, 23078 (see section 2.9) was probed with a radioactively-labeled $(A_3G)_{10}$ oligonucleotide probe to identify the $(A_3G)_n$ STRs (fig. 3.4).

The chromosome 19 recombinant cosmid clones will be referred to as chromosome 19 cosmids in the text. Three chromosome 19 cosmids 18618, 29395 and 20381 produced strong signals, chromosome 19 cosmids 20381, 19161 and 23078 gave fainter signals (fig. 3.4). The cosmids were prioritised for STR development based on the intensity of the signal, assuming that there is direct relationship between the number of $(A_3G)_n$ repeats and intensity. In addition, the position of the cosmids relative to the genetic map was also taken into account, thus cosmids 18618 and 29395 were selected and used for the development of $(A_3G)_n$ markers. The $(A_3G)_n$-containing fragments of these cosmids

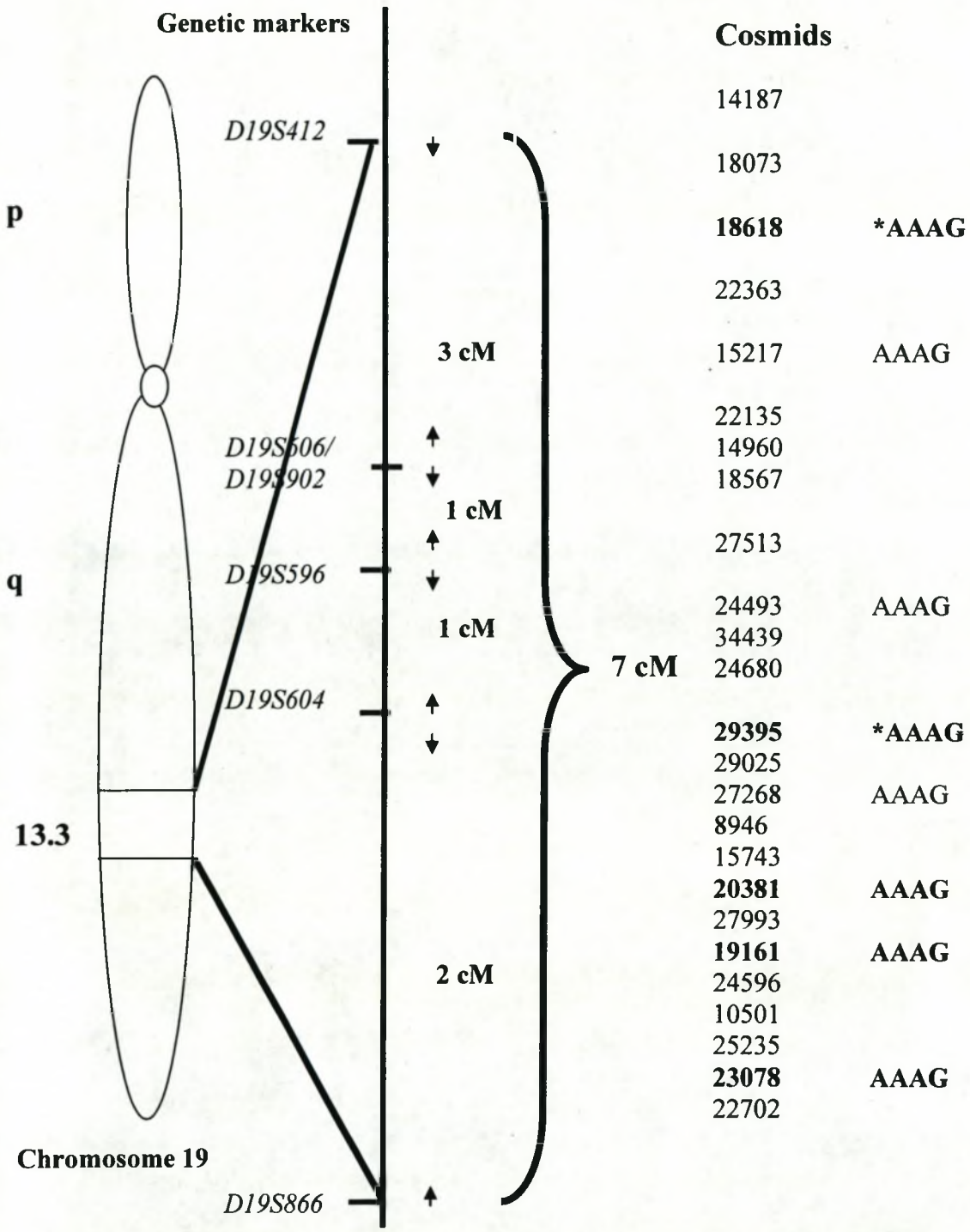


Figure 3.3 Schematic map of cosmid clones spanning the "start of study" 7cM *PFHBI* locus screened for $(A_3G)_n$ repeats.
Cosmids clones which gave a positive signal on probing with a $(A_3G)_{10}$ are in **bold.** *, The clones which were pursued for development as markers in this study (see fig. 3.4).

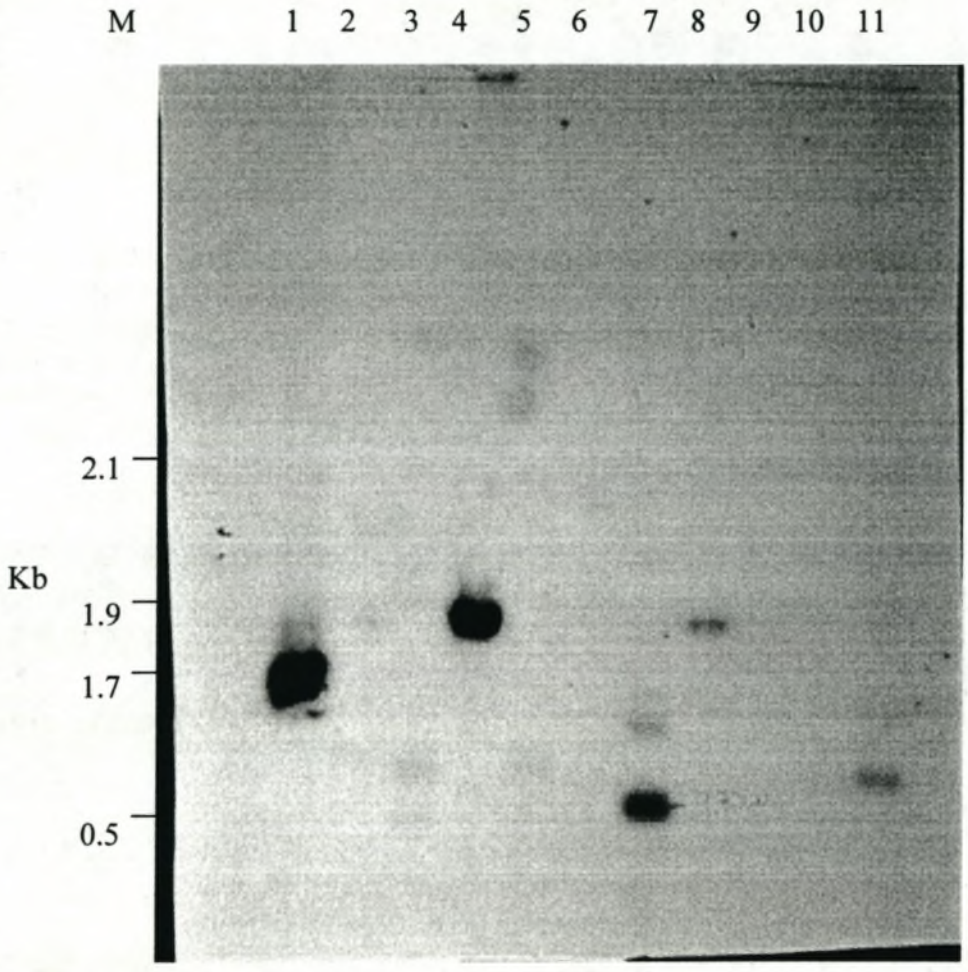


Figure 3.4 Identification of digested cosmid insert fragments harbouring $(A_3G)_n$ repeat motifs.
A Southern blot filter with *Sau*3A1-digested cosmid DNA probed with the [γ–^{32}P]dATP labeled $(A_3G)_{10}$ oligonucleotide.

Lane:	M: λPstI marker	6. 8946
	1. **18618**	7. 20381
	2. 15217	8. 19161
	3. 24493	9. 24596
	4. **29395**	10. 25235
	5. 27268	11. 23078

The cosmids in lanes 1 and 4 were selected for development as $(A_3G)_n$ markers

were subcloned into the plasmid vector pBluescript SK and used to transform *E. coli* strain XL1-blue as described in section 2.9. Colony blots, performed in duplicate, of the resulting recombinant plasmid transformants revealed several positive signals, indicating successful sub-cloning of the target fragment (fig. 3.5).

Cosmids 18618 and 29395

To confirm the subcloning of the $(A_3G)_n$ repeat motifs, recombinant plasmid DNA minipreps from positively hybridising colonies were digested with *Kpn*I and *Sac*I restriction endonucleases, located on either side of the multiple cloning site (fig. 2.4), to release the insert from clones 18618 and 29395, which were then sequenced. However, sequencing of the insert fragment released from cosmid 18618 did not reveal any $(A_3G)_n$ repeat motifs.

Sequencing of the insert fragment released from cosmid 29395 revealed an $(A_3G)_n$ repeat motif which comprised a string of 9 Gs, followed by 6 (A_3G) repeat motifs, 1 A_2G, 1 A_3G 2 Gs and 18 As (figs. 3.6 and 3.7). A BLAST homology search of the sequence containing the (A_3G) repeat motif and the flanking sequences revealed the presence of *Alu* sequences in the flanking sequences. The development of this marker was not pursued, since another member of the PFHBI project group was developing an $(A_3T)_n$ repeat marker from the insert of cosmid 29395 (Makubalo, 2000). In that study, an interrupted (A_3T) STR was

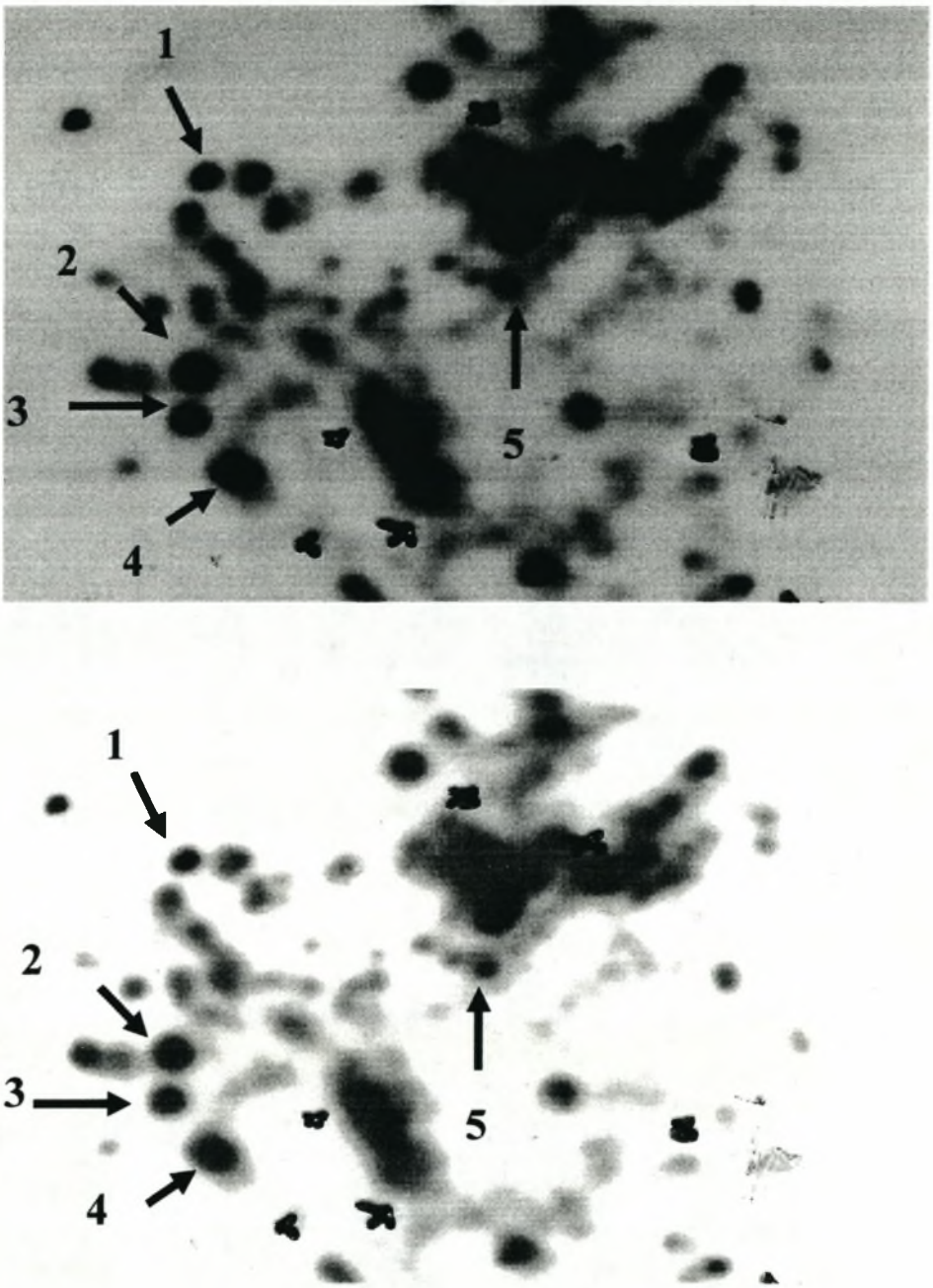


Figure 3.5 Autoradiograph of duplicate colony blot hybridisation of sub-cloned cosmid fragment 29395. The cosmid 29395 positively hybridising fragment was subcloned in Bluescript SK vector and the transformed colonies bound to the nylon membrane were probed with a labeled $(A_3G)_{10}$. The duplicate colonies giving strong signals were numbered as indicated by arrows 1 to 5. The orientation of the duplicate filters are shown as crosses on the autoradiographs.

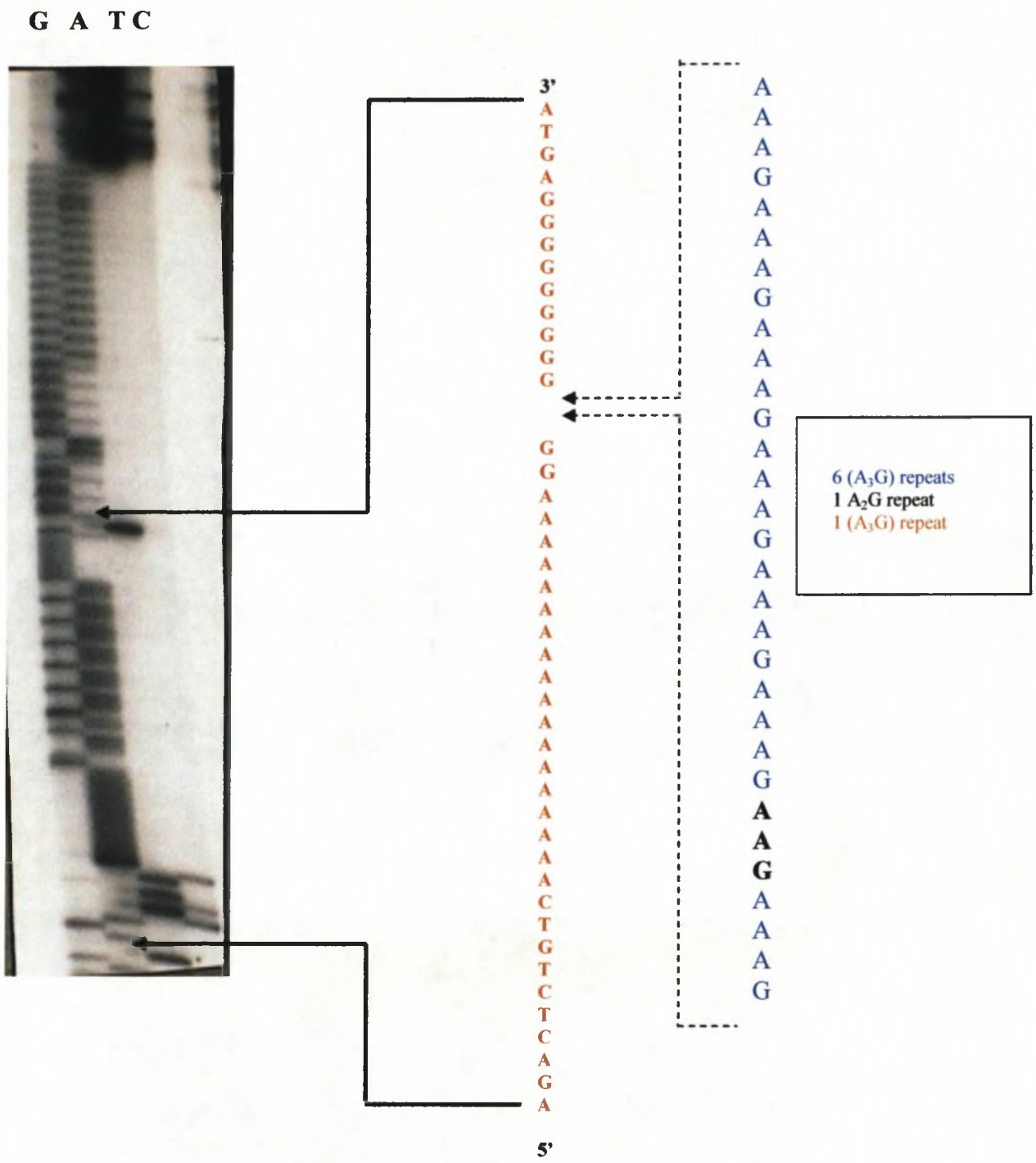


Figure 3.6 Autoradiograph of the partial sequence of the sub-cloned cosmid fragment 29395 sequences using pBluescript primer T7.
The sequence of the (AAAG) repeat motif is shown on the right with the dashed line indicating the position within the flanking sequences on the left. The sequence is interrupted by an **AAG** which shown in black. The start and end of the flanking sequence is indicated by the solid arrows on the autoradiograph. The sequence read is given in fig. 3.7.

5'-

AGACTCTGTCAAAAAAAAAAAAAAAAAGGGAAAGAAGAAAGAAAGAAA

GAAAGAAAGAAAGGGGGGGGGAGTAGGGAGGGAGGGAGAGAAGGAAG

GAGGAGGAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGA

AGGAAGGAAGGAAGG –3'

Figure 3.7 A portion of the sequence of the cloned cosmid insert 29395 harbouring the $(A_3G)_7$ repeat motif.
The flanking sequences is shown in orange. Part of the *Alu* sequence is underlined in red. The **$(A_3G)_7$ repeat motif** is interrupted by a **GAA.**

identified which comprised four runs of (A_3T) repeat motifs, followed by the sequence AATAAATAAT, which was followed by another run of (A_3T) repeat motifs. The $(A_3T)_n$ repeat marker was tested in a panel of control individuals but no polymorphic size variation was obtained (Makubalo, 2000).

3.1.4 Separation of *D19S606* and *D19S902*

As the refinement of the physical map of chromosome 19 proceeded, *D19S606* and *D19S902* were separated in May 2003 (fig. 3.8), placing *D19S902* telomeric to *D19S606*. BLAST homology searches against the human genome database allowed the physical distance between *D19S606* and *D19S902* to be determined. These searches places *D19S606* at 20,241,753bp and *D19S902* at 20,600,218bp on contig NT_011109 and therefore separated these two markers by a physical distance of 358,465bp (fig. 3.8). The genotyping results generated with *D19S902* for kindred 4 of pedigree 2 for the four

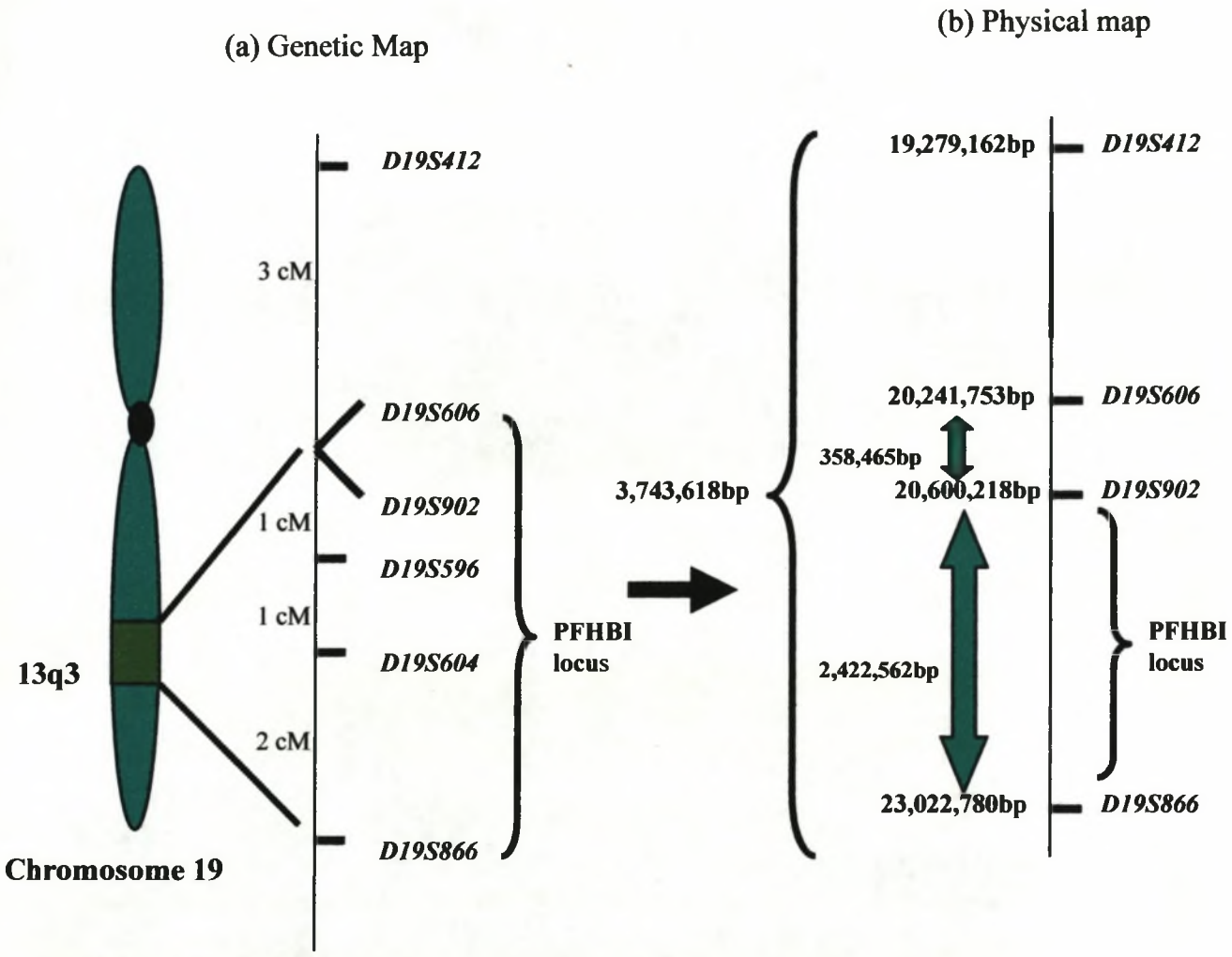


Figure 3.8 Maps showing the separation of genetic markers *D19S606* and *D19S902* (May 5, 2003). (a) A genetic map of the PFHBI locus showing the position of the genetic markers *D19S412, D19S606, D19S902, D19S596, D19S604* and *D19S866* in centimorgans (cM). In this map, *D19S606* and *D19S902* are extremely tightly linked. (b) A physical map of the region showing the separation of markers *D19S606* and *D19S902* by sequencing and the distances between them in bp.

PFHBI-affected individuals 4.III.2, 4.IV.1, 4.IV.2 and 4.V.2 showed that they carried allele 3 of marker *D19S902,* whereas, the three PFHBI-affected individuals from kindred 9 of pedigree 2 (9.III.3, 9.III.4 and 9.III.8) carried allele 5 of marker *D19S902* (fig. 3.9). These data refined the recombination breakpoint, resetting the centromeric limit at *D19S902,* and further reducing the PFHBI locus in physical terms by 358,465bp. The length of the physical distance of the PFHBI locus between genetic markers *D19S902* and *D19S866* was calculated to be 2,422,562bp (fig. 3.8).

3.1.5 Development of novel $(CA)_n$ repeat markers from database sequence

As the HGP progressed and published sequence data generated from cosmid and BACs became available, the natural progression was to develop STRs from sequence data rather than using cloning methods, for fine mapping of the PFHBI locus. In this study, selected $(CA)_n$ markers were developed from sequences deposited in publicly available databases in a further attempt to determine the breakpoint between markers D*19S902* and *D19S596*. Twenty seven BAC and cosmid sequences containing $(CA)_n$ repeat motifs (fig. 3.10) spanned the 2.42Mb of the PFHBI locus, between markers *D19S902* and *D19S866*. Of these, seven clone sequences (AC008392, AC008403, AC008888, AC010619, AC011495, AC018766 and AC010624) were selected for development as potential genetic markers (fig. 3.10). Clones AC008392, AC008403 and AC008888 were selected because they physically mapped to a region centromeric to *D19S596* and telomeric to *D19S902*, whilst the remaining four clones physically mapped to a region telomeric to marker *D19S604* and centromeric to *D19S866* (fig. 3.10).

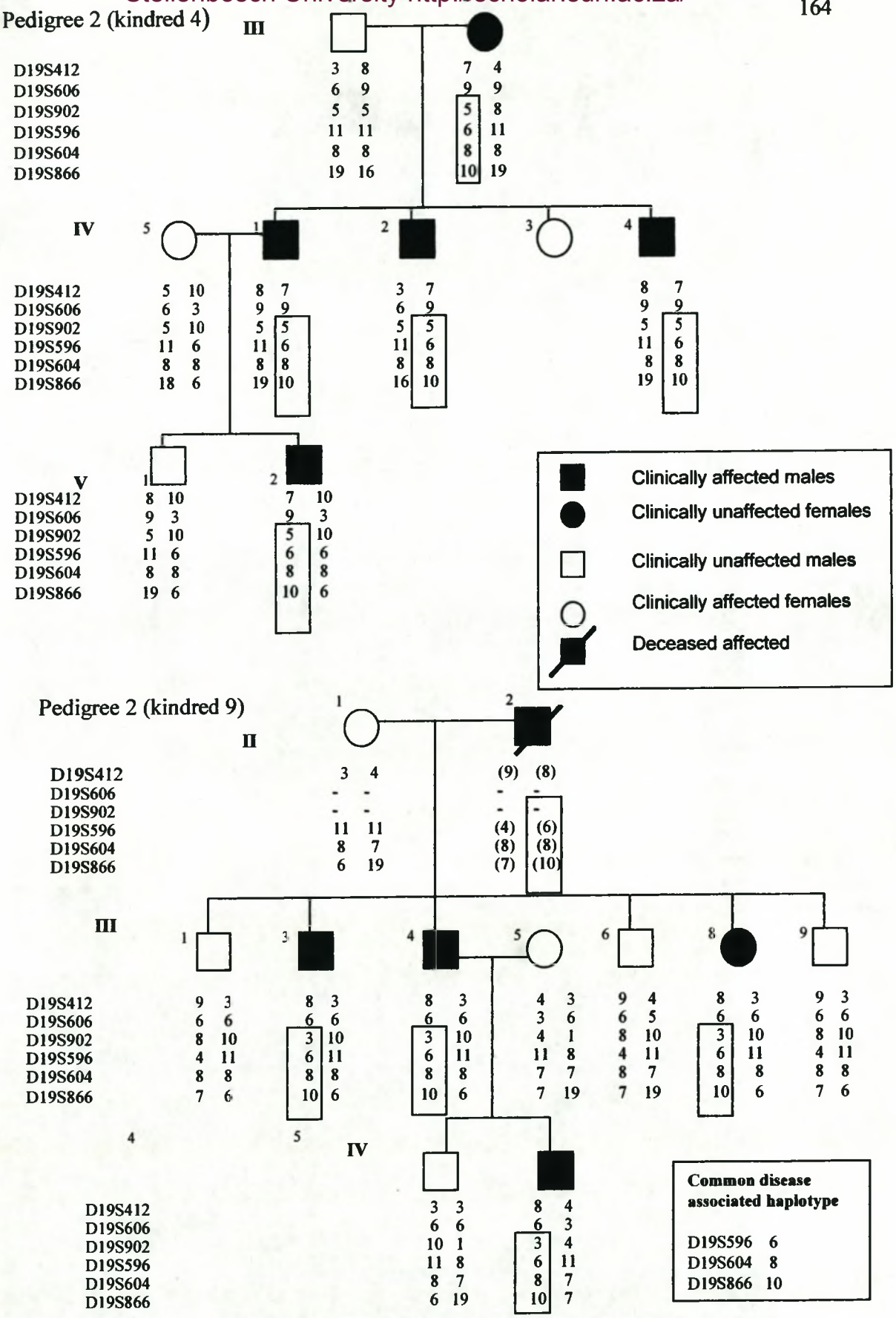


Figure 3.9 Genotyping with marker *D19S902* in kindreds 4 and 9 of pedigree 2.
Kindreds 4 and 9 are shown with the disease-associated haplotype at markers *D19S412, D19S606, D19S902, D19S596, D19S604* and *D19S866*. Haplotypes which have been inferred are indicated with brackets. The common disease-associated haplotype in each kindred is boxed. The pedigree generations are indicated by roman numerals.

Four clones (AC008392, AC008403, AC011495 and AC018766), out of the selected seven, were successfully developed as markers, which were subsequently submitted to Genbank and given the accession numbers *G68141*, *G68142, G68143* and *G68144*, respectively (table 3.1). The markers *G68141* and *G68142* were genotyped in PFHBI families and demonstrated a Mendelian pattern of inheritance, as shown (fig. 3.11a). The genotyping results obtained with marker *G68144* are not shown, since only one band was generated in all the samples.

The remaining three clones (AC00888, AC010619 and AC010624) were not used for marker development, for the following reasons. Firstly, the primers generated from the deposited sequence of clone AC008888 did not yield a PCR product with the clone AC00888, which suggests that the sequence in the database for AC00888 was incorrect. Secondly, the BLAST results of the sequence used for marker development did not correspond to clone AC010624 in the nucleic acid database, thus indicating that homology was not shared with the specified clone. This marker was not genotyped in the PFHBI families. Thirdly, 500bp of clone AC010619 mapped to two places on contig NT_011109, namely, to a position inside the PFHBI locus and to another telomeric to the PFHBI locus. Genotyping with this marker in the PFHBI families showed that it was unlinked, because the allele typing was different in each family (data not shown), suggesting that this clone might have been incorrectly placed and belonged elsewhere or that the 500bp sequence of this clone was duplicated on contig NT_011109.

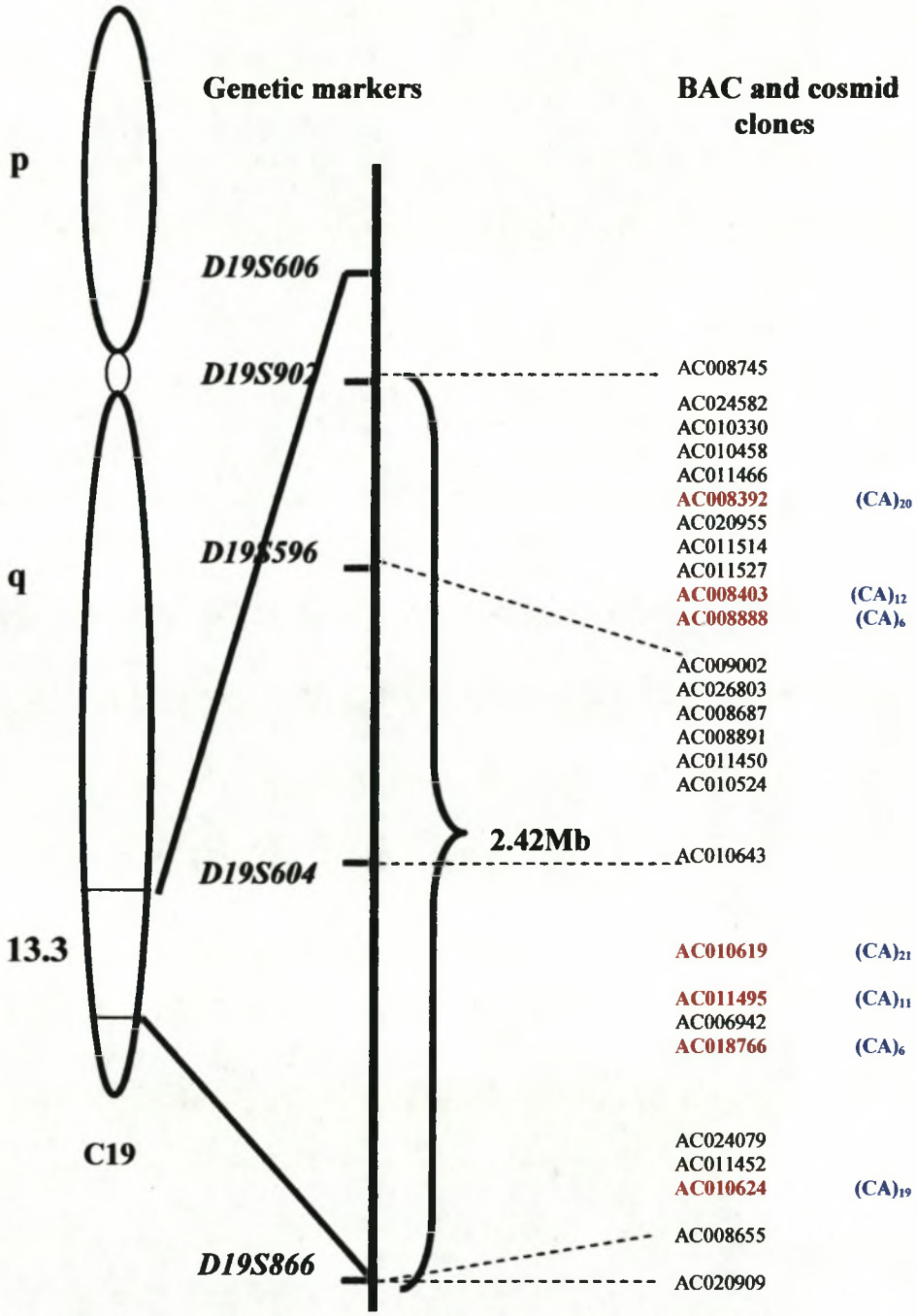


Figure 3.10 Schematic drawing of the PFHBI locus showing the relative positions of the BAC and cosmid clones containing $(CA)_n$ repeat motifs, identified by database searches.
Clones selected for the development of $(CA)_n$ repeat markers, as detected by database searches, are shown in red. The number of $(CA)_n$ repeat motifs contained in the selected clones according to database sequence are shown in blue. The positions of the available database genetic markers on the DNA clones are indicated by dotted lines.

Table 3.1 Primer sequences of novel $(CA)_n$ markers generated from recombinant BAC and cosmid sequences.

BAC/cosmid Clone Acc No	**STS Acc No**	**Primer sequence**	**T_a**	***Size of product**
AC008392	*G68141*	F 5'TAGGTTGGTCAGGCTTCGTG3' R 5'-CTCTGCTTTCTGGGTTCAAG-3'	56^{0}C	248bp
AC008888		F 5'-GTGGATCACCTGAGGTCAGG-3' R 5'-AGCTCCTGTTGCCCAGGCTG-3'	59^{0}C	226bp
AC008403	*G68142*	F 5'-GTGTTGGGATTGCAGGTGTG-3' R 5'-AGGCAAAGGATGCAGCGAAC-3'	59^{0}C	249bp
AC010619		F 5'-TAACATAAAGGGCTGTCTTC-3' R 5'-GTGAAAATAGAACAGCTA-3'	50^{0}C	169bp
AC011495	*G68143*	F 5'-TTGCAGTGAGCTGAGATCAC-3' R 5'-TCGCATCTAGAACCTATTGC-3'	55^{0}C	202bp
AC018766	*G68144*	F 5'-CTAATCCACTATGGCTGGTG-3' R 5'-GAAATCTTACAGGGGAAGGG-3'	55^{0}C	171bp
AC010624		F 5'-GGGTAAATGCTTGACTCTCC-3' R 5'-AGCCAGTTCCTGTCCTTATG-3'	58^{0}C	225bp

The clone accession numbers, the STS accession number, the primer sequences, optimum annealing temperatures and the expected amplicon size, estimated size from database sequence, for each marker are indicated. *, the Genbank accession numbers of the submitted STSs.

The allelic frequency of the three markers (*G68141*, *G68142* and *G68143)* was estimated by genotyping in an unrelated South African Afrikaner panel. The novel marker *G68141* was more polymorphic than markers *G68142* and *G68143* since the number of alleles for each marker was found to be 16, 4 and 2, respectively. Allele 12 of marker *G6841*

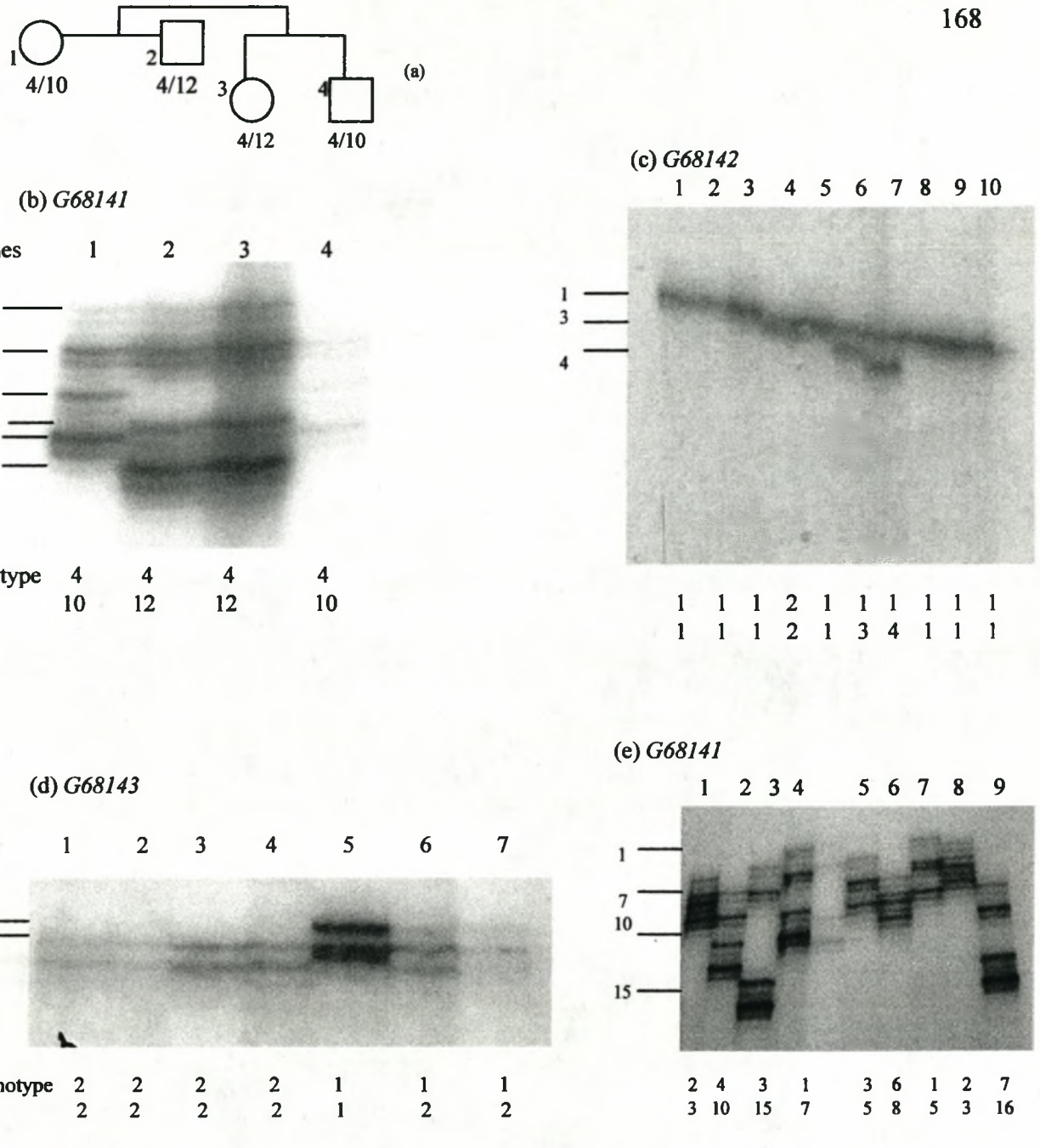


Figure 3.11 Autoradiographs of alleles of three polymorphic CA repeat markers developed from sequenced BAC and cosmid clones.
(a) A kindred consisting of parents (1 and 2) and children (3 and 4) of kindred 10 of pedigree 2 genotyped with (b) marker *G68141*. The allelic typing is shown in (b) and illustrates a Mendelian pattern of inheritance for marker *G68141*. The different alleles of markers typed with markers *G68142, G68143 and G68141* in DNA from PFHBI individuals correspond to (c), (d) and (e), respectively.

Table 3.2 Allele frequencies for the marker *G68141* in the South African Afrikaner population group

Allele number	Allelic frequencies
1	0.02
2	0.06
3	0.04
4	0.06
5	0.10
6	0.10
7	0.06
8	0.12
9	0.04
10	0.12
11	0.04
12	0.14
13	0.02
14	0.02
15	0.02
16	0.04

Number of individuals genotyped = 25

Table 3.3 Allele frequencies for the marker *G68142* in the South African Afrikaner population group

Allele number	Allelic frequencies
1	0.925
2	0.025
3	0.025
4	0.025

Number of individuals genotyped = 20

Table 3.4 Allele frequencies for the marker *G68143* in the South African Afrikaner population group

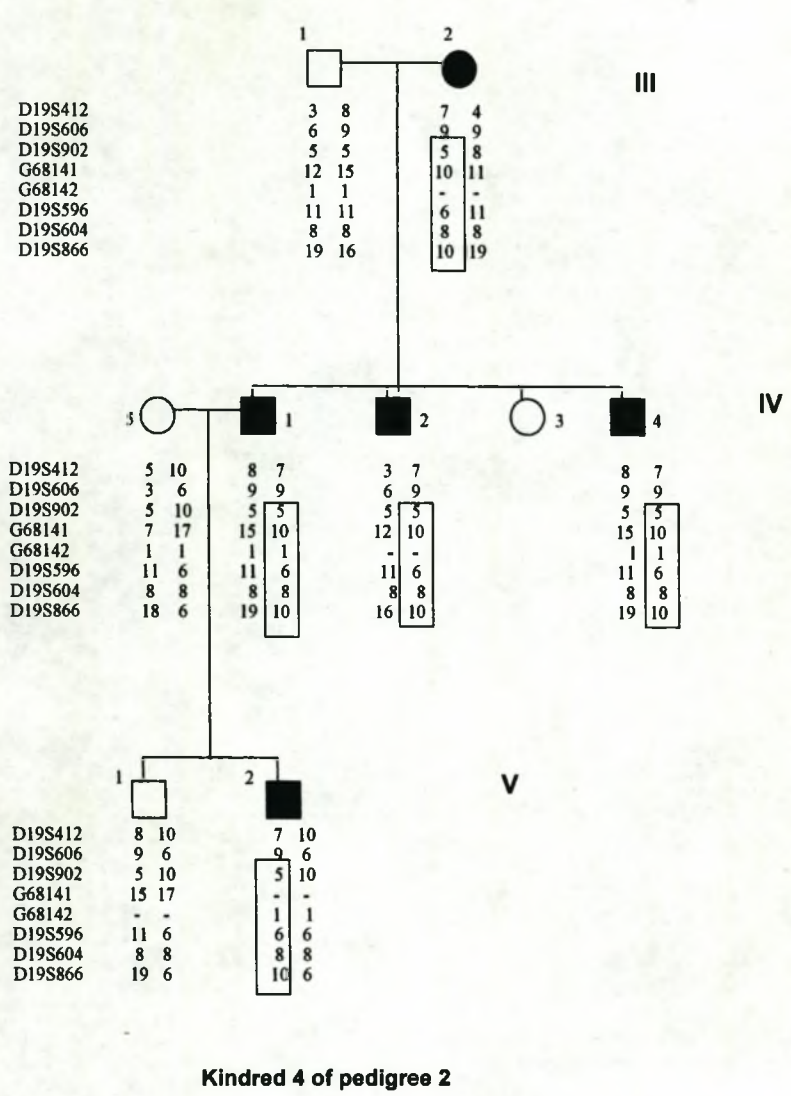
Allele number	Allelic frequencies
1	0.35
2	0.65

Number of individuals genotyped = 20

occurred most frequently with an allelic frequency of 0.14 (table 3.2), allele 1 of marker *G68142* with an allelic frequency of 0.925 (table 3.3) and marker *G68143* has two alleles with at an allelic frequency of 0.35 and 0.65 (table 3.4). Marker *G68144* was not polymorphic when typed in the unrelated South African Afrikaner population (results not shown). All four markers were submitted to the STS database. Genotyping results with markers *G68141*, *G68142* and *G68143* are shown in fig. 3.11 (c), (d) and (e).

Genotyping the PFHBI families with *G68141* and *G68142*

Since markers *G68141* and *G68142* lie centromeric to *D19S596* (fig. 3.10) and novel markers were required from this region in order to further refine the breakpoint in kindred 4 of pedigree 2, both *G68141* and *G68142* were used to genotype individuals from kindred 4 of pedigree 2. Individuals from kindred 9 of pedigree 2 were also genotyped with both markers as a reference (fig. 3.12). The genotype for the disease-associated haplotype (D19S902-*G68141-G68142-D19S596-D19S604-D19S866)* was 5-10-1-6-8-10 in kindred 4 (4.III.2, 4.IV.1, 4.IV.2, 4.IV.4 and 4.V.2) and 3-10-1-6-8-10 in kindred 9 (9.III.3, 9.III.4 and 9.III.8 and 9.IV.5), which showed that the centromeric breakpoint lies between *D19S902* and *G68141*. This result suggests that more markers should be developed from sequence between *D19S902* and *G68141*, in order to see how close the breakpoint lies to *D19S902*. The physical distance between *G68141* and *D19S902* was estimated to be 447kb.



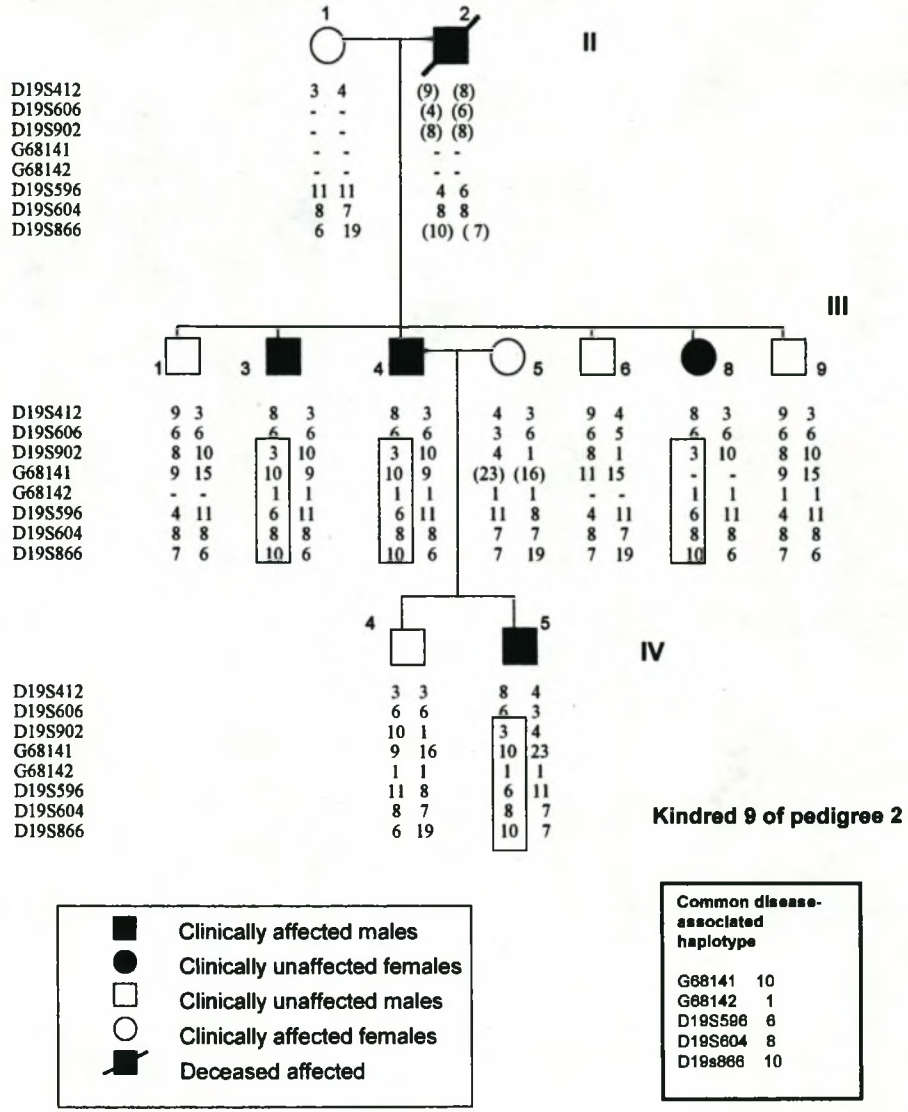


Figure 3.12 Genotyping kindreds 4 and 9 of pedigree 2 with novel $(CA)_n$ markers, *G68141* and *G68142*.
Kindreds 4 and 9 are shown with the disease-associated haplotype at markers *D19S412, D19S606, D19S596, D19S604, D19S866* and novel markers *G68141* and *G68142* . The disease-associated haplotype in each kindred is boxed. Haplotypes, which have been inferred, are indicated with brackets. The pedigree generations are indicated by roman numerals. -, Samples in which markers did not generate any products and could therefore not be genotyped.

3.2 DATABASE SEARCHES TO IDENTIFY GENE TRANSCRIPTS

In June 1999, a search of the EST databases Genemap99 and the LLNL site was undertaken to find ESTs which mapped within the PFHBI locus, 38 transcripts were identified (table 3.5). Each EST was extended using STACKpack from the STACK database resulting in the generation of a number of STACK clusters, e.g., EST M78727 yielded one STACK cluster, while EST M85740 yielded 85 STACK clusters (see table 3.5). Two ESTs (G13834 and G17554) had no cluster hits. However, protein homology searches using all these clusters yielded only six proteins, namely, CA11, two zinc finger proteins, a hypothetical protein, collagen related KIAA089 and a myc protein, which corresponded to ESTs AF067662, M85740, G26380, T97152, AA031950 and F02213, respectively (table 3.5) (only protein homologies with the highest scoring BLAST values were selected) (June 1999). At this stage, very little information was available about these proteins and other gene identification and selection strategies were followed.

However, in May 2003, the same 38 ESTs were clustered using the UNIGENE database (NCBI) and the protein homology searches were repeated (table 3.6). Fifteen ESTs matched proteins in the nonredundant (nr) databases shown in red (table 3.6), of

Table 3.5 STACK clusters and protein homologies of ESTs retrieved from Genemap99 and the LLNL site (June 1999).

EST	Number of STACK clusters	Protein homology
M78727	1	
AFO67662	10	Carbonic anhydrase
M14269	2	
*M85740	85	Zinc finger protein
T17236	4	
T47573	6	
T77818	1	
H48539	1	
R11661	2	
*T74222	76	
N25336	5	
R09427	2	
R26303	6	
R46185	2	
AA057647	7	
R61534	2	
G13826	2	
Z39055	7	
T90778	6	
G13834	no cluster hits	
*R11034	12	
G15328	2	
T87340	6	
T02911	1	
*T66137	17	
G26380	14	Hypothetical protein
*T85363	9	
Z41308	4	
R37469	2	
R49148	9	
G22461	6	
H50984	2	
*R79413	6	
*T97152	82	Zinc finger protein
G17554	no cluster hits	
AA031950	15	Collagen related KIAA089
Z39409	6	
F02213	15	Myc protein

EST, expressed sequence tag. * Indicates EST expression in heart. The number of STACK clusters was obtained by doing a BLAST search against the STACK database.

which six were expressed in cardiac tissue namely, KPTN, CA11, CRD8, PLEKHA4 ZNF473 (alias ZFP100) and ATF5, indicated by an asterisk (*). An example of redundancy of the EST databases was observed since three different ESTs (M78727, T17236 and Z41308) had homology to the same protein KPTN.

Nine ESTs (T77818, R26303, R61534, R11043, T87340, T66137, G26380, R37469 and F02213) had homology to a number of different predicted proteins, of which three proteins (FLJ2040, FLJ25129fis and FLJ33742fis) were expressed in heart (indicated by a cross (+) in table 3.6). Six ESTs (M14269, R11661, G13826, G13834, G15328 and G17554) had homology to regions outside the PFHBI locus or to another chromosome, (shown in blue in table 3.6).

Two ESTs (T47573 and R11661) had homology to *Alu* sequences, one within and one outside the PFHBI locus, respectively. Twelve ESTs (M14269, H48539, R09427, AA057647, G13826, G13834, G15328, T02911, G22461, H50984, R79413 and G17554) had no homology to any protein.

The retrieved ESTs were mapped to three contigs, NT_011190, NT_011140 and NT_011157, spanning the PFHBI locus (fig. 3.13) using their physical (base pair) position (table 3.6) and later to the merged contig NT_011109 (fig. 3.15). Annotated genes *KPTN*, *ZNF473* and *ATF5* from table 3.6 were identified as the most important

Table 3.6 Protein homologies and the base pair positions of ESTs on the contig NT_011109 (May 2003).

EST	Acc No.	Protein homology	Position on contig NT_011109 (bp)
M78727	M78727	*KPTN	20,248,249bp
AFO67662	XM_055856	*CA11/CARP2	20,246,994bp
M14269	M14269	none	chromosome 7
M85740	XM_009313	ZNF264	20,695,400bp
T17236	XM_055856	KPTN	20,248,249bp
T47573	T47573	Alu sequence	20 118,554bp
T77818	XM_091974	LOC147891	20,292,145bp
H48539	H48539	none	19,699, 700bp
R11661	R11661	Alu 5 sequence	NT_011255 on chromosome 19
T74222	NM_005500	SAE1	19,980,632bp
N25336	XM_085943	GRLF1	19,772,900bp
R09427		none	19,855,157bp
R26303	NM_017805	+FLJ20401	21,493,023bp
R46185	XM_038970	SLC8A2	20,200,108bp
AA057647	AA057647	none	19,792,906bp
R61534	AK057858	+FLJ25129 fis	20,267,776bp
G13826		none	chromosome 11
Z39055	NM_031485	GRWD	21,225,353bp
T90778	XM_049938	*CRD8/TUCAN	20,974,604bp
G13834		none	chromosome 6
R11034	XM_055859	MGC45806	20,514,581bp
G15328	NM_017805	none	chromosome 2
T87340		FLJ20401	21,492,032bp
T02911		none	20,241,639bp
T66137	XM_058466	FLJ20643	22,217,746bp
G26380		+FLJ33742fis	21,787,341bp
T85363		*PLEKHA4	21,608,544bp
Z41308	XM_055856	KPTN	20,246,598bp
R37469	NM_020309	cDNAFLJ33742fis	22,200,853bp
R49148	XM_097340	VRK3	22,747,915bp
G22461		none	21,983,283bp
H50984	138022	none	22,077,971bp
R79413		none	
T97152		*ZNF473	22,818,615bp
G17554		none	chromosome 10
AA031950		AP2A1	22,578,558bp
Z39409	NM_012068	*ATF5	22,705,264bp
F02213	NM_032712	Hypothetical protein MGC13170	23,569,152bp

Legend of table 3.6

*/+, Indicates expression in heart tissue. *, annotated gene; +, predicted gene. ESTs matching proteins in the nonredundant database are shown in **red.** The ESTs M14269, M85740, R11661, G13826, G13834, G15328 and G17554, shown in **blue,** map outside the PFHBI locus. The proteins underlined in red are most highly prioritised ones.
KPTN, actin-binding protein; *CA11/CARP2*, carbonic anhydrase XI; *ZNF264*, zinc finger 264; *LOC147891*; hypothetical protein LOC147891;*SAE1*, sumo activating enzyme subunit 1; *GRLF1*, glucocorticoid receptor DNA binding factor 1; *FLJ20401*, hypothetical protein FLJ20401; *SLC8A2*, *Homo sapiens* solute carrier family 8 (sodium-dependent inorganic phosphate cotransporter), member 8; FLJ25129, hypothetical protein FLJ25129; GRWD, *Homo sapiens* glutamate rich WD repeat protein; *CRD8/TUCAN*, tumor up-regulated CARD-containing antagonist of caspase nine; *FLJ20643*, hypothetical protein FLJ20643; *SLC17A7*, *Homo sapiens* solute carrier family 17 (sodium-dependent inorganic phosphate cotransporter), member 17; *VRK3*, vaccinia-related kinase 3; *I38022*, hypothetical protein I138022; *ZNF473*, zinc finger protein 473; *ATF5*, activating transcription factor 5; *MGC13170*, hypothetical protein MGC13170.

candidate genes, based on their position, expression in cardiac tissue and their involvement in cardiac function, as plausible PFHBI candidate genes. The genes obtained from this search were also mapped to the integrated map (section 3.3, fig. 3.15) and were included in the evaluation process as potential PFHBI candidates (section 3.4).

3.3 A HIGH RESOLUTION MAP

In order to facilitate the search for the PFHBI gene, a high resolution map of the PFHBI locus, using the methods described in section 2.12, was generated. In February 2002, three contigs NT_011190, NT_011140 and NT_011157 spanned the PFHBI locus (fig. 3.13). In May 2003, the three contigs were merged into one contig, NT_011109. An

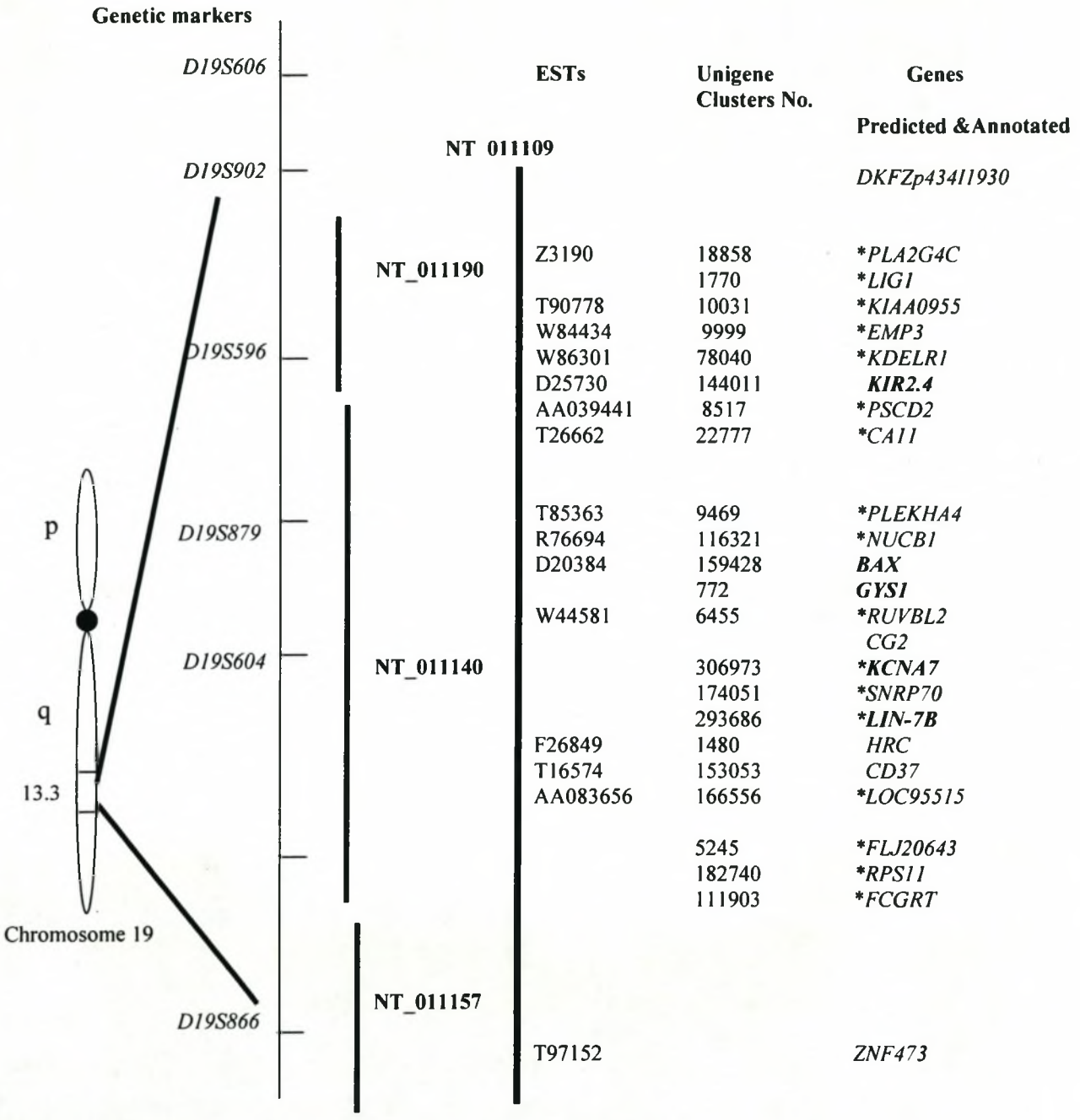


Figure 3.13 An early map of the three contigs spanning the PFHBI locus (February 5, 2002).
From left to right: Diagrammatic representation of chromosome 19 showing the position of the PFHBI locus, genetic map, 3 contigs NT_011190, NT_011140 and NT_011157 and the merged contig NT_011109, the positions of retrieved ESTs, EST clusters and some of the genes att the PFHBI locus *, genes showing expression in heart and other tissues. The genes in **bold** were genes selected for PFHBI mutation screening for this study. The gene abbreviations are:
DKFZp434I1930, predicted gene; *PLA2G4C*, phospholipase A2, group IVC (cytosolic, calcium-independent); *LIG1*, ligase I DNA, ATP-dependent; *KIAA0955*, predicted gene; *EMP3* epithelial membrane protein 3; *KDELR*, (Lys-Asp-Glu-Leu) endoplasmic reticulum protein retention receptor 1; *KIR2.4*, potassium inwardly-rectifying channel, subfamily J, member 1; *PSCD2*, pleckstrin homology, sec7 and coiled-coil domain 2; *CA11* carbonic anhydrase X1; *PLEKHA4*, pleckstrin homology domain containing, family A (phosphoinositide binding specific) member 4; *NUCB1*, nucleobindin 1; *BAX*, *BCL2*-associated X protein; *GYS1*, glycogen synthase 1; *RUVBL2* RuvB-like 2 *(E. coli)*; *CGB2*, chorionic gonadotrophin beta polypeptide 2; *KCNA7*, potassium voltage-gated channel, shaker-related subfamily, member 7; *SNRP70*, small nuclear ribonucleoprotein 70kDa polypeptide (RNP antigen); *LIN-7B*, lin-7 homolog B (*C. elegans*); *HRC*, histidine rich calcium binding protein ; *CD37*, CD37 antigen; *LOC95515*, predicted gene; *FLJ20643*, hypothetical gene; *RPS11*, ribosomal protein S11; *FCGRT*, Fc fragment of IgG, receptor, transporter, alpha; *ZNF473*, zinc finger protein 473.

updated map of the PFHBI locus was generated using the sequence data from contig NT_011109 (fig. 3.15).

3.3.1 Experimental data

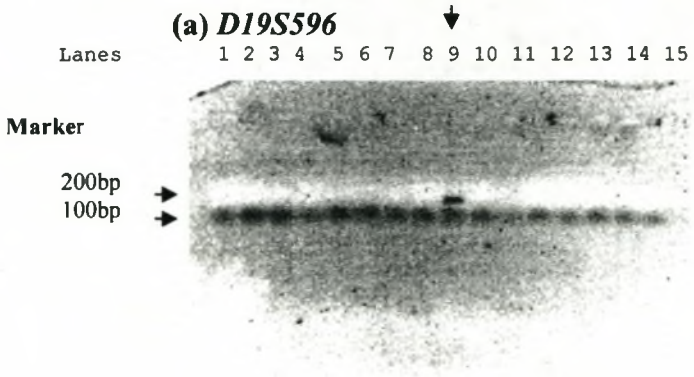
PCR amplification using primers sets of genetic markers *D19S902*, *D19S606*, *D19S596*, *D19S604*, *D19S879* and *D19S866* with BAC and cosmid DNA clones spanning the PFHBI area were performed (section 2.12.2). A representative data set of the PCR amplification results is shown (fig. 3.14). PCR products were obtained with markers *D19S596*, *D19S879* and *D19S866* using BAC clones AC009002, AC008687 and AC020906, respectively (fig. 3.14 and table 3.7).

Subsequent computational searches, which involved a BLAST search of both published and novel genetic markers (section 2.12.1) against the human genome database, was performed.

3.3.2 Bioinformatic data

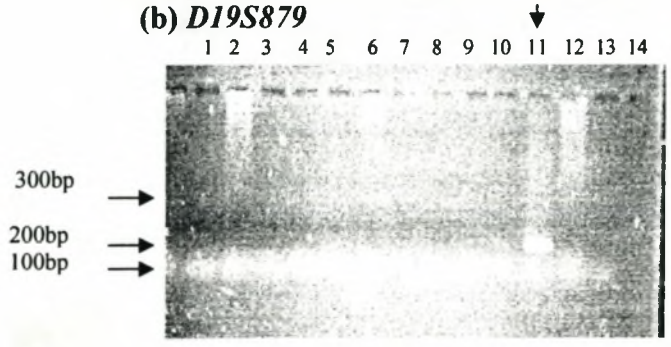
NT_011109 contig

Homology search results positioning the database genetic markers on the corresponding BAC clones and the NT_011109 contig are shown in table 3.8, while table 3.9 contains the BLAST results of the novel $(CA)_n$ repeat markers from section 3.1.5. BLAST results generated for the retrieved 38 ESTs (section 3.2) are shown in table 3.6, indicating the position of the 38 ESTS on contig NT_011109 spanning the PFHBI locus. In table 3.10,



Lanes:

1. AC008745	9. **AC009002**
2. AC024582	10. AC0115727
3. AC010330	11. **AC008687**
4. AC010458	12. **AC020906**
5. AC011466	13. AC008403
6. AC008392	14. AC008888
7. AC020955	15. AC026803
8. AC011514	



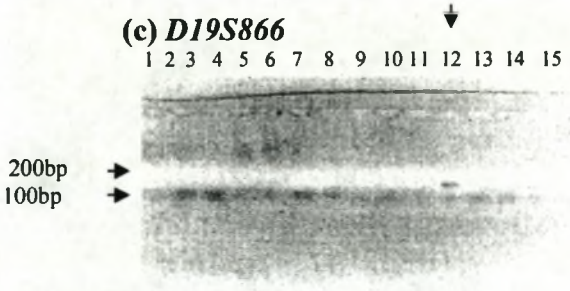


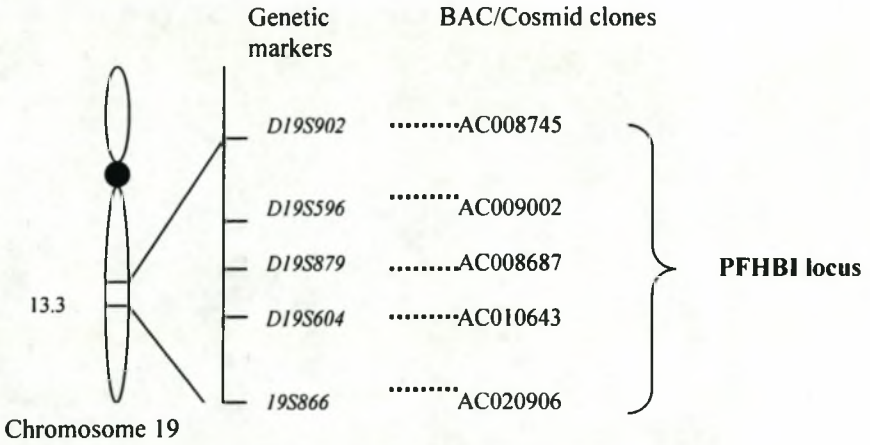
Figure 3.14 PCR amplification products obtained using primers of genetic markers *D19S596*, *D19S879* and *D19S866* and DNA clones spanning the PFHBI locus as templates.
(a) PCR product obtained with genetic marker *D19S596* observed in lane 9. (b) PCR product obtained with genetic marker *D19S879* observed in lane 11. (c) PCR product obtained with genetic marker *D19S866* observed in lane 12. The position of the various 100bp marker bands is shown on the left hand side of the gels. ↓ shows the lane in which the PCR product occurs (see table 3.7).

(a) **Table 3.7 Position of genetic markers on DNA BAC clones using PCR amplification**

		Genetic Markers				
BAC or cosmid clones		***D19S902***	***D19S596***	***D19S879***	***D19S604***	***D19S866***
Genbank Acc No	**LLNL ID**					
AC008745	**BC782556**	++	-	-	-	-
AC024582	BC858854		-	-	-	-
AC010330			-	-	-	-
AC010458	BC694129		-	-	-	-
AC011466	BC324323		-	-	-	-
AC008392	BC242886		-	-	-	-
AC020955	R33773		-	-	-	-
AC011514	F16353		-	-	-	-
AC009002	**R31763**		++	-	-	-
AC011527			-	-	-	-
AC008687	**BC52309**		-	++	-	-
AC010643	**R29295**		-	-	++	
AC008403	BC255070		-	-	-	-
AC008888	BC008888		-	-	-	-
AC026803	BC808483		-	-	-	-
AC020906	**BC61330**		-	-		++

(b)

(b)



(a) PCR amplification results obtained with *D19S902*, *D19S596*, *D19S879*, *D19S604* and *D19S866* primer sets using template DNA clones spanning the PFHBI locus. (b) Schematic representation of the results from (a). LLNL ID, Lawrence Livermore National Laboratories identity number for the chromosome 19 BAC and cosmid clones; Genbank Acc No., accession number for each DNA clone; ++, PCR product present; -, no PCR product. Horizontal dotted line indicates marker identified with the corresponding clone.

Table 3.8 Position of genetic markers on contig NT_011109 (May 5, 2003)

Genetic Markers	Genbank Acc No	BAC clone Acc No	BLAST results bp
D19S902	Z53165	AC008745	20600218–20600469
D19S596	Z52219	AC009002	21519195-21519452
D19S879	Z52767	AC008687	21784108-21784383
D19S604	Z53032	AC010643	22165196–22164953
D19S866	Z52072	AC008655 AC020909	23022780–23022989

The base pair position of the genetic marker on contig NT_011109 and the DNA clone to which the genetic marker has homology, is shown. Genbank Acc No., genbank accession number for each genetic marker; bp, base position.

Table 3.9 Position of novel $(CA)_n$ markers on contig NT_011109 (May 5, 2003)

New $(CA)_n$ Markers Genbank Acc No	BAC clone Acc No	BLAST results bp
G68141	AC008392	21046497-21046744
G68142	AC008403	21292264-21292512
G68143	AC011495	22406103-22406304
G68144	AC018766	22598888-22599058

The base pair position of the genetic marker on contig NT_011109 and the DNA clone, to which the genetic marker has homology is shown. Genbank Acc No., genbank accession number for each novel genetic marker; bp, base position

Table 3.10 The position of annotated genes on contig NT_011109 at the PFHBI locus (order from centromere to telomere – *D19S902* to *D19S866*)

Gene	Acc No	BAC Acc No	Bit score	E value	Position of gene on contig NT_011109 (bp)	Size of BAC/cosmid bp	Position of BAC/cosmid on contig NT_011109	Gap size between non-overlapping BACs (bp)
KPTN	NM_007059	AC010331	754	0	20246994–20248249	80240	20293689-20373929	117443
D19S902	Z53165	AC008745			20600218-20600469	180531	20491372-20671903	30369
CABP5	NM_019855	AC010458	1215	0	20802037-20815494	105788	20702272-20809060	44797
PLA2G4C	NM_003706	AC010458	7031	0	20819836-20881933			
LIG1	NM_000234	AC011466	558	e^{-156}	20887182-20928591	167272	20853857-21021129	3881
CARD8	NM_014959	AC011466	512	e^{-142}	20983157-21005925			
EMP3	NM_001425	AC008392	469	0	21101748-21098336	91842	21025010-21116852	
		AC020955				41589	21085318-21126907	
KDELR1	XM_006801	AC011514	635	e^{-179}	21154441-21162949	39379	21124301-21163680	
GRWD	NM_031485	AC011527	560	e^{-157}	21217453-21224472	35438	21159820-21195258	
PSCD2	NM_004228	AC008403	314	e^{-82}	2120797-21245464	190076	21187945-21378021	131354
KIR2.4	NM_170720	AC008403	3728	0	21235619- 21233885			
SULT2B	NM_004605	AC008403	1037	0	21346987-21370872			
CA11	NM_0021217		1488	0	21417443 – 21411540			
FGF21	NM_019113	AC009002	562	e^{-157}	21527684-21528476	37402	21509375-21546777	
PLEKHA4	NM_020904	AC026803	998	0	21640048-21609461	224271	21534113-21758384	1755
NUCB1	NM_006184	AC026803	1834	0	21693762-21672231			
BAX	NM_004324	AC026803	875	0	21732255-21726410			
GYS1	NM_002103	AC026803	2805	0	21764721-21739625			
RUVBL2	NM_006666	AC008687	281	e^{-73}	21775721-21786050	157633	21760139-21917772	95623
KCNA7	NM_031886	AC008687	6657	0	21844328-21838865			
SNRP70	NM_003089	AC008687	1252	0	21857860-21879890			
LIN-7B	NM_022165	AC008687	410	e^{-112}	21887708-21889907			
HRC	NM_002152	AC008891	3846	0	21926854-21922648	117628	22013395-22131023	
CGB7	NM_001774	AC011450	569	e^{-160}	22111696-22106925	106168	22101102-22207270	
CD37	NM_001774	AC011450	242	e^{-128}	22106988-22111775			
TEAD2	NM_003598	AC011450	1627	0	22112042-22133904			

Table 3.10 cont

SLC17A7	AB032436	AC011450	500	e^{-138}	22212766-22202002				
		AC10524				37349	22110286-22147635		
D19S604	Z53032	AC010643			22165196-22164953	38902	22145764-22184666		
FCGRT	NM_004107	AC010619	629	e^{-177}	22297454-22284728	179394	22184276-22363670		
RRAS	NM_006270	AC011495	706	0	22407108-22411393	209844	22345634-22555478		
SRA-1	AF254411	AC011495	685	e^{-150}	22416474-22434609				
BCL2L12	NM_052842	AC011495	569	e^{-149}	22437271-22440555				
AP2A1	NM_014203	AC006942	429	e^{-159}	22553392-22577700	42710	22555829-22598539		
		AC018766				41624	22595942-22637566	27738	
AKT1S1	NM_032375	AC024079	446	e^{-122}	22644742-22641364				
TBC1D17	NM_024682	AC024079	404	e^{-109}	22649171-22660071				
ATF5	NM_012068	AC011452	2909	0	22700650-22705381	146741	22665304-22812045		
VRK3	NM_016440	AC011452	589	e^{-165}	22748236-2277901				
ZNF473	XM_046390	AC011452	7891	0	22816116-30224513				
		AC10624				167787	22763297-22931084	121094	
		AC008655				123149	23052178-23175327		
D19S866	Z52072	AC020909			23022780-23022989	100826	23173141-23273967		
		AC010515				41518	23019563-23061081		

Acc No is the genbank accession number. The bit score means the number of the bases matching the query sequence with the nucleic acid sequence in the atabase. E value is the expectation value and is treated like probability values. E values of smaller than 0.05 or 0.02 are considered as significant. Genetic markers *D19S902, D19S604* and *D19S866* of the *PFHBI* locus are shown in underlined.

Gene abbreviations are: *KPTN*, actin binding protein; *CABP5*, calcium binding protein 5; *PLA2G4C*, phospholipase A2, group IVC (cytosolic, calcium-indep ndent); CARD8, caspase recruitment domain family member 8;*EMP3* epithelial membrane protein 3; *KDELR*, (Lys-Asp-Glu-Leu) endoplasmic reticulum protein retention receptor 1; *GRWD*, glutamate rich WD repeat protein; *PSCD2*, pleckstrin homology, sec 7 and coiled-c il domain 2; *KIR2.4*, potassium inwardly-rectifying channel, subfamily J, member 1; *SULT2B*, sulfotransferase family, cytosolic, 2B, member 1; *CA11* carbonic anhydrase X1; *FGF21*, fibroblast growth factor 21; *PLEKHA4*, pleckstrin homology domain containing, family A (p hosphoinositide binding specific) member 4; *NUCB1*, nucleobindin 1; *BAX*, *BCL2*-associated X protein; *GYS1*, glycogen synthase 1; *RUVBL2* RuvB-like 2 *(E. coli)*; *KCNA7*, potassium voltage-gated channel, shaker-related subfamily, member 7; *SNRP70*, small nuclear ribonucl oprotein 70kDa polypeptide (RNP antigen); *LIN-7B*, lin-7 homolog B (C. elegans); *HRC*, histidine rich calcium binding protein ; *CD37*, CD37 antigen; *CGB7*, chorionic gonadotrophin beta polypeptide 7; *TEAD2*, TEA domain family member 2; *FCGR*, Fc fragment of IgG, eceptor, transporter, alpha ; *SLC17A7*, solute carrier family 17 (sodium independent inorganic phosphate cotransporter, member 7; *RRAS*, related RAS viral oncogene homolog *BCL2L12*. *BCL2*-like 12 (proline rich); *AP2A1*, adaptor-related protein complex 2, alpha 1 subunit *AKT1S1*, AKT1 substrate 1 (proline-rich); *TBC1D17*, TBC1 domain family, member 17; *ATF5*, activating transcription factor 5; *VRK3*, vaccinia related kinase 3; *ZNF473*, Zinc finger protein 473.

the homology search results generated for annotated genes have been placed at the PFHBI locus with their corresponding BAC clones.

Using the updated bioinformatics data (May 2003), an integrated map of the PFHBI locus was generated (fig. 3.15). However, there were some discrepancies with map data from table 3.10. Retrieved clone sequence AC018766 could not be positioned on the contig NT_011109 because there was not enough overlap. The gene density per clone differed, e.g., only *HRC* was present on BAC clone AC008891, whereas AC026803 contained four genes, *PLEKHA4*, *NUCB1*, *BAX* and *GSY1* and, in another case, *CA11* which had not been assigned to a BAC but placed directly onto contig NT_011109 (table 3.10).

In all cases, the bioinformatic results supported the experimental results, with the exception of genetic marker *D19S866*, which has BLAST homology to clones AC008655 and AC020906, whereas the experimental data placed this marker only on DNA clone AC020906. Some possible explanations could be that part of the sequence or the total insert containing the marker was lost during isolation of the BAC DNA clone or that the PCR was not successful.

3.4 EVALUATION OF GENES AS POTENTIAL CANDIDATES

An investigation of the merged contig, NT_011109 (fig. 3.15), showed that 82 genes were present at the PFHBI locus and therefore several criteria were used to prioritise the

annotated and predicted genes, as described in section 2.13. The results of the analyses are shown in tables 3.11, 3.12 and 3.13.

3.4.1 Annotated and predicted genes on contig NT_011109

Of the total of 82 genes placed at the PFHBI locus, 35 annotated genes were present on contig NT_011109 (table 3.11, fig. 3.15). Twenty six genes are expressed in heart tissue, indicated, in turn, in bold in table 3.11, making them plausible candidates for PFHBI and they should, in turn, be further investigated by mutation screening. Genes that did not show cardiac expression was most probably due to the absence of data rather than the absence of expression. Predicted proteins were analysed by the identification of protein domains, which have specific functions that are similar to known proteins. These proteins were therefore prioritised based, mainly, on the importance of the predicted function of their identified domains in cardiac physiology. Forty-seven predicted genes were present on contig NT_011109, of which only six genes, indicated by an asterisk (*) in table 3.12, have predicted functional domains, which it could be speculated, may, if mutated, affect cardiac function and therefore cause PFHBI.

3.4.2 Prioritised candidate genes

A list of plausible candidate genes was generated (table 3.13) using the criteria for prioritisation described in section 2.13.1. The reason for prioritising genes were that a large number of plausible PFHBI candidate genes were present at the locus therefore, genes were prioritised ranging from A, the most highly prioritised, to D, the least

Table 3.11 Annotated genes at the PFHBI locus on contig NT_011109

GENES	*EXPRESSION*	*FUNCTION*
D19S902		
CABP5	retina	belongs to subfamily of calcium binding proteins EF-hand–signaling proteins and buffering/transport proteins eg calmodulin calcium-binding motif
PLA2G4C	multiple+ **heart**	cytosolic calcium independent phospholipase A2
LIG1	multiple+ **heart**	functions in DNA replication and the base excision repair process
CARD8	multiple+ **heart**	Cancer associated gene. Inhibits NF-kappa-B activation
EMP3	multiple+ **heart**	involved in cell proliferation and cell-cell interactions
KDELR1	multiple+ **heart**	endoplasmic reticulum protein receptor, protein has been isolated from the golgi apparatusglutamate receptor, receptor family-ligand binding region, ligand-gated ion channel region
GRWD	multiple	WD domain, WD40 repeat domain
PSCD2	multiple+ **heart**	cytosolic guanine nucleotide-exchange factor for ADP-ribosylation factor ARF6
KIR2.4	multiple	potassium inward-rectifying channel
SULT2B1	multiple+ **heart**	sulfotransferase domain
CA11	multiple+ **heart**	belongs to a large metallo enzymes. Plays a general role in the central nervous system.
FGF21	multiple + **heart**	Branched chain aminotransferase 2, mitochondrial
PLEKHA4	multiple+ **heart**	fibroblast growth factor 21
NUCB1	multiple+ **heart**	secreted protein with DNA binding activity, has signal peptide

GENES	EXPRESSION	FUNCTION
BAX	multiple	involved with the regulation of apoptosis
GYS1	muscle+**heart**	glycogen metabolism, transfers glycosyl residue from UDP-glucose to glycogen
RUVBL2	multiple + **heart**	DNAB-like helicase domain, similar to bacterial recombination factor *E. coli* RuvB, ATP dependent DNA helicase, chaperone, protein complex assembly, multi-chaperone pathway, DNA repair
KCNA7	amnion_normal, muscle+**heart**	potassium channel protein
SNRP70	multiple+ **heart**	pre-mRNA splicing factor, RNA binding mRNA processing
LIN-7B	multiple+ **heart**	PSD-95 domain, PDZ domain (also known as DHR or GLGF, a possible role is the anchoring of ion channels at specific subcellular sites
HRC	ear, kidney, muscle + **heart**	histidine–rich calcium-binding protein
CD37	multiple+ **heart**	cell surface glycoprotein, transmembrane 4 family, predicted to have 4 transmembrane spanning domain
CG7	multiple	function unknown
TEAD2	multiple + **heart**	Transcription factor. Expressed during early embryonic development
FCGRT	multiple+ **heart**	Fc fragment of IgG, receptor transporter, alpha
D19S604		
SLC17A7	multiple	Transports glutamate into synaptic vesicles.
RRAS	multiple + **heart**	R-Ras promotes tumor growth of cervical epithelial cells and increases their migration potential over collagen through a pathway that involves PI 3-K.
SR-A1	multiple + **heart**	predicted splicing factors
BCL2L12	multiple	involved in the regulation of apoptosis
AP2A1	multiple + **heart**	involved in golgi to endosome transport, protein transporter activity
AKT1S1	multiple	function unknown
TBC1D17	multiple	involved in the signal transduction mechanism
ATF5	multiple + **heart**	Basic-leucine zipper (bZIP) transcription factor. Binds the camp response element (cre) (consensus
VRK3	multiple + **heart**	Regulation of transcription
ZNF473	multiple + **heart**	Regulation of transcription
D19S866		

Published genetic markers *D19S902*, *D19S604* and *D19S866* are in bold and underlined. Gene abbreviations are: *CABP5*, calcium binding protein 5; *PLA2G4C*, phospholipase A2, group IVC (cytosolic, calcium-independent); CARD8, caspase recruitment domain family member 8; *EMP3* epithelial membrane protein 3; *KDELR*, (Lys-Asp-Glu-Leu) endoplasmic reticulum protein retention receptor 1; *GRWD*, glutamate rich WD repeat protein; *PSCD2*, pleckstrin homology, sec 7 and coiled-coil domain; *KIR2.4*, potassium inwardly-rectifying channel, subfamily J, member 1; *SULT2B*, sulfotransferase family, cytosolic, 2B, member 1; *CA11* carbonic anhydrase X1; *FGF21*, fibroblast growth factor 21; *PLEKHA4*, pleckstrin homology domain containing, family A (phosphoinositide binding specific) member 4; *NUCB1*, nucleobindin 1; *BAX*, *BCL2*-associated X protein; *GYS1*, glycogen synthase 1; *RUVBL2*. RuvB-like 2 *(E. coli)*; *KCNA7*, potassium voltage-gated channel, shaker-related subfamily, member 7; *SNRP70*, small nuclear ribonucleoprotein 70kDa polypeptide (RNP antigen); *LIN-7B*, lin-7 homolog B (C. elegans); *HRC*, histidine rich calcium binding protein ; *CD37*, CD37 antigen; *CGB7*, chorionic gonadotrophin beta polypeptide 7; *TEAD2*, TEA domain family member 2; *FCGR*, Fc fragment of IgG, receptor, transporter, alpha ; *SLC17A7, solute carrier family 17 (sodium independent inorganic phosphate cotransporter, member 7; RRAS*, related RAS viral oncogene homolog; *SRA-1*, serine arginine-rich pre-mRNA splicing factor; *BCL2L12*, *BCL2*-like 12 (proline rich); *AP2A1*, adaptor-related protein complex 2, alpha 1 subunit; *AKT1S1*, AKT1 substrate 1 (proline-

Table 3.12 Predicted genes on contig NT_011109

PREDICTED GENE	ACC NO.	FUNCTIONAL DOMAIN
LOC147892	XP_097342	No function assigned
LOC147891	XP_091074	Myb-like DNA-binding receptor binding N-coR and TFIIIB DNA-binding CpG island domain
LOC126155	XP_064953	periplasmic substrate binding
LOC126152	XP_064952	FAR1 FAMILY
LOC163071	XP_091973	krueppel associated box similar to HSPC059
LOC147890	XP_085954	No function assigned
LOC93233	XP_049939	No function assigned
**FLJ10922*	XP_012808	61% protein similarity to alpa –1C- andrenergic receptor splice form 2, 40% protein similarity to *Caenorhabditis elegans* collagen 2, tyrosine kinase catalytic domain
KIAA1883	XP_055866	No function assigned
LOC147888	XP_085935	No function assigned
LOC147887	XP_055867	similar to ribosomal protein L18
**LOC126147*	XP_058987	laminin EGF-like domain netrin C-terminal domain glycosyl transferase domain
LOC93236	Not found	No function assigned
LOC147886	Not found	No function assigned
LOC93237	XP_049980	glycosyl transferase family similar to FUT2 in function
LOC126145	Not found	No function assigned
LOC147885	XP_085938	ras association domain
**FLJ10922*	Not found	402 aa similar to collagen alpha 2(I) domain
**FLJ20200*	Not found	apoptosis protein-caspase recruitment protein, has a DIL domain, may be a myosin heavy chain
KIAA0955	Not found	No function assigned
FLJ20401	Not found	No function assigned
LOC163070	XP_091972	ER lumen protein retaining receptor
LOC147886	XP_009103	glycosyl transferase family
LOC147884	XP_009100	fibroblast growth factor region
LOC51171	Not found	retinal short chain dehydrogenase
LOC147883	XP_097346	No function assigned
LOC93314	XP_050460	No function assigned
LOC147882	XP_10323	No function assigned
LOC147881	XP_085957	No function assigned
LOC147880	XP_097328	No function assigned
LOC147879	XP_097345	No function assigned
LOC147875	XP_097331	No function assigned
LOC147874	Not found	No function assigned
FLJ10490	Not found	No function assigned
LOC163067	XP_103238	No function assigned
LOC147873	XP_097341	No function assigned
**LOC147872*	XP_085941	myosin tail ezrin/radixin/moesin family
LOC113091	XP_053723	similar to tuberofundibular 39
LOC57333	NP_065701	No function assigned
LOC147871	XP_085948	No function assigned
LOC126133	XP_058991	similar to aldehyde dehydrogenase
LOC51171/	Not found	No function assigned
FLJ10490	Not found	No function assigned
**LOC95515*	XP_027924	similar to TEA domain family member 2.
FLJ20643	XP_027903	No function assigned
KIAAA1205	XP-046305	structural constituent of cell wall
FLJ226888	NP_079405	function unknown

LOC, Locuslink sequences; *, genes shown in the prioritised list in table 3.13 (section 3.4)
FLJ, from the NEDO human cDNA sequencing project of the Kazusa DNA Research Institute;
KIAA, from the Kazusa cDNA sequencing project of the Kazusa DNA Research Institute);
XP, Genbank protein sequences

prioritised based on their role in cardiac function. The most highly prioritised three genes selected to be screened in category A were genes which are involved in ion channel function and recruitment, namely, *KCNA7*, *KIR2.4* and *LIN-7B*, since mutations in the ion channel gene *SCN5A* and *Kir2.1* cause several cardiac disorders accompanying conduction disturbance, such as PCCD (Schott *et al.*, 1999) and Andersons syndrome (Plaster *et al.*, 2001), respectively (sections 1.3.2 and 1.5.2.2).

Category B lists two calcium binding genes, *HRC* and *CAPB5,* the rationale for selection of these two genes being that a mutation in a Ca^{2+}-binding protein would result in an ion imbalance, possibly resulting in an electrophysiological disturbance causing arrhythmia (section 1.6.3) (Katz, 1993).

Category C lists *NUCB1* as a plausible candidate gene for a number of reasons. The rationale for the selection of the human *NUCB1* gene for PFHBI mutation screening was that studies showed that lupus-prone mice secrete a murine *Nucb1* protein that promotes the production of DNA-specific antibodies (Kanai *et al.*, 1993). The human homolog of these antibodies were also shown to be present in babies of mothers with systemic lupus erythromatosis (SLE) and one of the clinical symptoms of SLE is heart block in the absence of muscle failure (Lanham *et al.*, 1983). In addition, a previous study using PFHBI-affected individuals was undertaken to investigate the involvement of specific circulating autoantibodies in the pathology of the disease (Moolman, 1992). Additional

Table 3.13 Prioritised list of PFHBI candidates genes

CATEGORY	GENE	EXPRESSION	FUNCTION
A			
Ion channel genes	**KCNA7*	m+heart	ion channel function
	**KIR2.4*	m	ion channel function
	**LIN-7B*	m+heart	recruitment of ion channel
B			
Calcium-binding genes	*+HRC*	m+heart	calcium-binding protein
	CAPB5	m	calcium-binding calcium binding motif & transporting of proteins
C			
Regulatory genes	*+NUCB1*	m+heart	signal peptide,
	+SNRP70	m+heart	mRNA processing
	**/+GSY1*	m+heart	glycogen metabolism
	**BAX*	m	regulation of apoptosis
	EMP3	m+heart	cell proliferation control differentiation and death
	TEAD2	m	transcriptional activity
	ZNF473	m+heart	transcriptional activity
	CD37	m+heart	transport and development
D			
Predicted genes	FLJ10922	-	tyrosine kinase domain
	LOC126147	-	lamin EGF-like domain
	FLJ10922	-	collagen domain
	FLJ20200	-	apoptotic, myosin domain
	LOC147872	-	myosin tail
	LOC95515	-	transcription activity

*, genes selected for mutation screening and screened in this study. +, genes selected for mutation screening in previous studies. */+, partial screening in a previous study and the remaining screening were performed in this study. m, expression in multiple tissues excluding heart. Categories of priority ranges from A to D, where A is the most, and D is the least, prioritised for mutation screening.

reasons for selecting *NUCB1* were that it has triplet repeat in exon 13 and has transcriptional factor activity (F. February, 2002).

A gene involved in development, namely, a splicing factor gene, *SNRP70,* was listed in category C. The rationale followed was that defective splicing of mRNA can be disease-causative (Faustino and Cooper, 2003), thus implicating splicing factor genes which, in turn, are involved in the development of organs and systems, possibly the cardiac conduction system. Therefore, a mutation in any gene involved in the splicing mechanism could be PFHBI-causative. Previous studies have shown that a mutation in the *PRPC*8 gene, which is a splicing factor pre-mRNA, caused RP13 in affected families (sections 1.6.1 and 1.6.3) (McKie *et al.*, 2001).

GSY1, in category C, was listed as a plausible candidate gene to screen because it has been shown experimentally that glycolytic flux maintains membrane activity, reduces arrhythmia and improves the functional recovery of the ischaemic heart (King and Opie, 1998). Therefore, a mutation in *GSY1* may affect the glycolytic levels, which may, in turn, affect the functioning of the heart and could possibly be PFHBI disease-causing. Recent findings have shown that GSK-3 β inhibits glycogen synthesis through phosphorylation of GSY1. *GSK-3 β* is involved in the regulation of heart development via the *Wnt* signaling pathway (Hardt and Sadoshima, 2002), which may also link *GSY1* as a role player in the complex pathway of cardiac development (section 1.5.4), thus further supporting its candidature as a PFHBI-causative gene.

Two genes involved in apoptosis, cell differentiation and proliferation namely, *BAX* and *EMP3*, were also listed in category C, because, in the human heart, the morphogenesis of the CCS and the normal postnatal involution of the right ventricle are both mediated by apoptosis. The exact mechanisms by which these events are initiated or terminated remain poorly understood (James *et al.*, 1996). Apoptosis is essential for normal development, excessive apoptosis may be triggered by pathological changes and results in the destruction of tissues and the development of heart disease in which a fatal arrhythmic event may occur (section 1.5.1) The reason for selecting genes with transcriptional activity, namely, *ZNF473* and *TEAD2,* as possible PFHBI candidates is that it has been shown that faulty transcription factors can cause congenital cardiac disorders, for example mutations in *TBX5* and *NKX2-5* can cause HOS and ASD, respectively (section 1.3.1) (Basson *et al.*, 1997; Schott *et al.*, 1998). Another plausible PFHBI candidate gene is *CD37*, which encodes a cell surface glycoprotein, predicted to have four transmembrane spanning domains, which appears to play a role in the development, regulation and proliferation of cell tissues. The expression of this gene in heart strengthens its PFHBI candidacy.

Category D lists genes of lowest priority, as it contains only predicted genes which have signature domains of genes which have been implicated as possible PFHBI candidates. For example, *FLJ10922* has a tyrosine kinase-like domain (table 3.13), since a mutation encoding a serine threonine kinase causes DM, which manifests cardiac conduction

abnormalities as one of its features (section 1.3.2.1) (Harper, 1989; Perloff *et al.*, 1984), it is plausible that a mutation in a *DMK* homolog may cause PFHBI.

Another predicted gene, *LOC126147,* has a signature with a lamin-like domain, and since mutation in a *LMNA* gene causes cardiac abnormalities, including A-V conduction disturbances in LGMD1B (section 1.3.2.2)((Kass *et al.*, 1994), one can propose that a *LMNA* paralog would be a plausible gene to screen for PFHBI-causative mutations. The three predicted genes, *FLJ10922, FLJ20200* and *LOC147872*, are plausible candidates as they have either a collagen or myosin signature, or both, respectively (table 3.13), since a defect in a structural gene expressed in the CCS, such as collagen or myosin, may be pathogenic. In addition, the increase in the collagen content during development coincides with the progressive course of PFHBI (section 1.2.3.1) (Brink *et al.*, 1994).

Predicted gene, *FL20200,* also has an apoptotic signature domain (table 3.13). While apoptosis is essential for normal development, if a gene involved in the apoptotic pathway is mutated, it can result in the triggering of excessive apoptosis and result in the destruction of tissues and the development of heart disease, such as, PFHBI in which a complete heart block may occur (section 1.5.1) (James *et al.*, 1996). Finally, predicted gene *LOC95515* has a signature domain which possesses transcriptional activity, since faulty transcription factors cause congenital cardiac disorders, such as, HOS and ASD (Basson *et al.*, 1997; Schott *et al.*, 1998), a mutation in this gene may possibly be PFHBI-causative.

3.5 FURTHER GENE SEARCHES AT THE PFHBI LOCUS

A search of the PFHBI locus for three candidate genes, namely, a *Cx* gene, a G-protein-encoding gene and a gene with a CTG repeat motif, none of which were known to map there, but which were strong functional candidates, was undertaken (sections 2.14.1 and 2.14.2). The results of these investigations are shown in sections 3.5.1, 3.5.2 and 3.5.3.

3.5.1 Connexin

The PCR amplification using the *Cx* degenerate primers (described in section 2.14.3) did not generate any products from the BAC and cosmid clones spanning the PFHBI locus; however, a product of approximate 176bp in size was obtained from two radiation hybrids, namely, RHG12 and ORIM-7, and from genomic DNA (fig. 3.16). The nucleotide sequence of this product is shown in fig. 3.17(a). In addition, a BLAST search of this sequence against the nr database showed that it had homology to a *Cx37* gene-encoding a *Homo(H) sapiens* gap junction protein, alpha 4, GJA4 (Acc no. NM_002060) (January 2004), with an E value of $4e.^{-13}$ The *Cx* PCR product sequence was matched to the partial sequence of *Cx37* mRNA (Acc. no. BC027889), from base pair position 139 to position 318 (fig. 3.17c). The BLAST2 alignment results of the sequence of the amplified product with each BAC or cosmid clone sequence spanning the PFHBI locus showed no homology with the 176bp *Cx* PCR product. A BLAST search of the sequence of the *Cx* PCR product against the contig NT_011109 and the clones spanning the PFHBI locus did not yield any homologies. At time of this analysis, the sequencing of chromosome 19 was still incomplete and a gap were still present in the contig spanning the PFHBI area

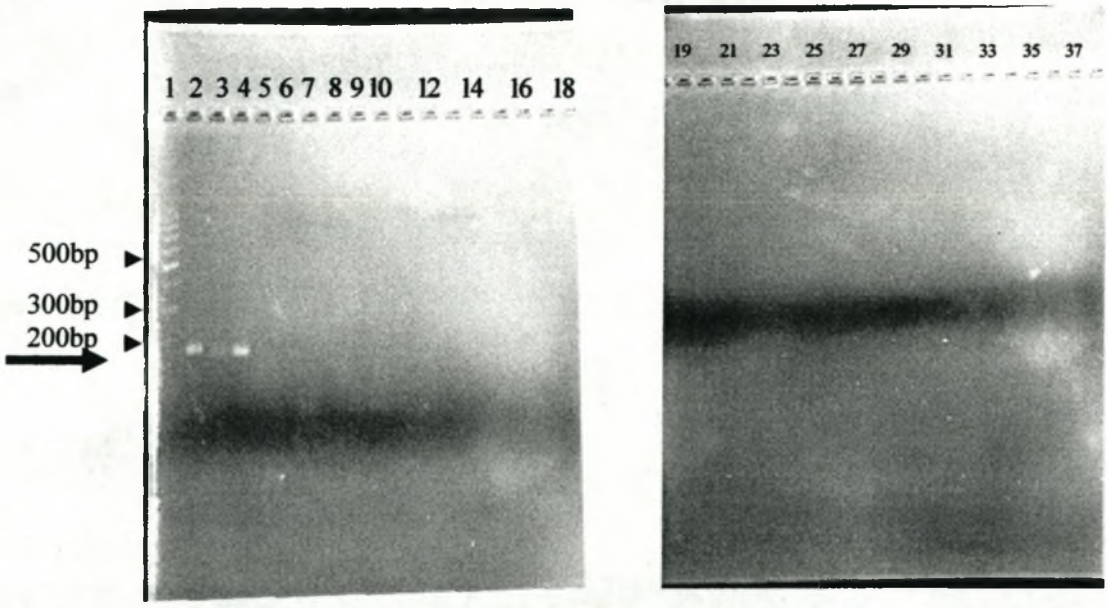


Figure 3.16 PCR amplified products obtained using *Cx* degenerate primers and DNA clones spanning the PFHBI locus.

Lanes:			
	1. 100 bp marker	13. BC694629	25. BC679592
	2. **genomic DNA**	14. BC324323	26. BC266129
	3. **RHG12**	15. BC242886	27. R31681
	4. **RH Orim-7**	16. R33773	28. R29295
	5. water control	17. F16353	29. BC878087
	6 BC894691	18. R26730	30. BC42053
	7. BC821616	19. BC255070	31. R31181
	8. F24003	20. R29279	32. F23669
	9. R30005	21. BC677569	33. R28785
	10. BC782556	22. R31763	34. R28785
	11. BC858854	23. BC808483	35. BC275645
	12. BC815354	24. BC52309	36. BC61330

The samples from which a PCR product was genetated (lanes 2, 3 and 4) are shown in **bold.** The position of the *Cx* PCR-amplified product is indicated by an arrow. lane 1, positive control; lane 5, negative control; RH, radiation hybrid; BC, bacterial artificial chromosome; R and F, cosmids.

(a)

```
GNNTNNNTTTNNNTTTCCTTTGGGCCANNNCATNNTGNCTGNACCTAGCT    50
GCTGANTNAGTGTGGGGCNGANGAGCAGTCTGATTTNGANTGNAACACTCT  100
CCAGCCAGGCTGCACACAACGTGTTGCTACGACCAGGCNTTCCCTATCTC   150
CCATATNCGCTACTGGGTGCTGCAGA                           176
```

(b)

Score = 110 bits (57), Expect = 9e-22
Identities = 103/123 (83%), Gaps = 3/123 (2%)
Strand = Plus / Plus

```
Query:       54  gantnagtgtggggcngangagcagtctgatttngantgnaacactctccagccaggctg 113
                 || | |||||||||  || ||||||||  |||||  ||  ||  |||||   ||||||||||||
Sbjct:       199 gagtcagtgtgggg-tgacgagcagtcagatttcgagtgtaacacggcccagccaggctg 257
GJA4 protein 42   E  S  V  W  G   D  E  Q  S  D  F  E  C  N  T  A  Q  P  G  C

Query:       114 cacacaacgtgttgctacgaccaggcnttccctatctcccatatncgctactgggtgctg 173
                 |||  ||||||  |  ||||  ||||||||  |||||  ||||||||  ||  ||||||||||||||
Sbjct:       258 cac-caacgtct-gctatgaccaggccttccccatctcccacatccgctactgggtgctg 315
GJA4 protein 62    T   N  V   C  Y  D  Q  A  F  P  I  S  H  I  R  Y  W  V  L

Query:       174 cag 176
                 |||
Sbjct:       316 cag 318
GJA4 protein 81   Q
```

(c)

```
  1 gcgggcgtca ctccggccat cgtccccacc tccacctggg ccgcccggca ggcaggcgac
 61 ggaggcccgg gagccatggg tgactggggc ttcctggaga agttgctgga ccaggtccag
121 gagcactcga ccgtggtggg taagatctgg ctgacggtgc tcttcatctt ccgcatcctc
181 atcctgggcc tggccggcga gtcagtgtgg ggtgacgagc agtcagattt cgagtgtaac
241 acggcccagc caggctgcac caacgtctgc tatgaccagg ccttccccat ctcccacatc
301 cgctactggg tgctgcagtt cctcttcgtc agcacaccca ccctggtcta cctgggccat
361 gtcatttacc tgtctcggcg agaagagcgg ctgcggcaga aggaggggga gctgcgggca
421 ctgccggcca aggacccaca ggtggagcgg gcgctggcgg ccgtagagcg tcagatggcc
```

Figure 3.17 Analysis of PCR amplified product obtained with *Cx* degenerate primers.
(a) The sequence of the *Cx* PCR product obtained with the radiation hybrid RG12 DNA sample. The sequence of the *Cx* PCR underlined in black matches the *Cx37* sequence encoding the GJA4 protein (b) BLAST 2 search of the sequence of the *Cx* PCR product (query) against *Cx37* (subject) (c) Partial sequence of *Cx37* mRNA (accession no. BC027889). The sequences matching the Cx degenerate **forward primer** are indicated in **blue** and are underlined and the sequences matching the Cx degenerate **reverse primer** are indicated in **red** and is underlined. Sequence of the *Cx* PCR amplified product matching *Cx37* mRNA is shown in bold.
E, glu; S, ser; V, val; W, Trp; G, gly ;D, asp; E, glu; Q, gln; F, phe; N, asn;T, thr; A, ala; P, pro; C, cys; T, thr; Y, tyr; I, ile; H, his; R, arg; L, leu.

(IHGSC, 2001), thus, there is still a possibility that a *Cx* might map to a gap at the PFHBI locus or to an area on chromosome 19q13.1 to 13.3. The sequence of the *Cx* PCR product also had homology to a rat connexin (results not shown) also with an E value of $4e,^{-13}$ which suggests that the connexin detected may have been from the hamster DNA of the chromosome 19 radiation hybrid. Unfortunately, rodent DNA was not included as a control in the PCR amplification so as to verify the BLAST result.

3.5.2 G protein

A dot blot of cosmids and chromosome 19 radiation hybrids spanning the PFHBI region was probed with radioactively labeled G-protein (*p115-RhoGEF* cDNA) (2.14.2) and the hybridisation results are shown in fig. 3.18. After stringent washing of the labeled dot blot, only the chromosome 19 radiation hybrids and the genomic DNA control gave positive signals. In addition, a BLAST2 search of *RhoGEF* cDNA sequence with each of the DNA clones spanning the PFHBI locus found no homologous sequence present on the clones. Since the chromosome 19 radiation hybrids spanned a region from chromosome 19q13.1 to 19q13.3 (fig. 1.10), the dot blot results suggest that this gene is present in this greater region. A subsequent bioinformatic search for this gene revealed that it mapped to contig NT_011109 in a region centromeric to the PFHBI locus within chromosome 19q13.1 to 19q13.1 region (fig. 3.19).

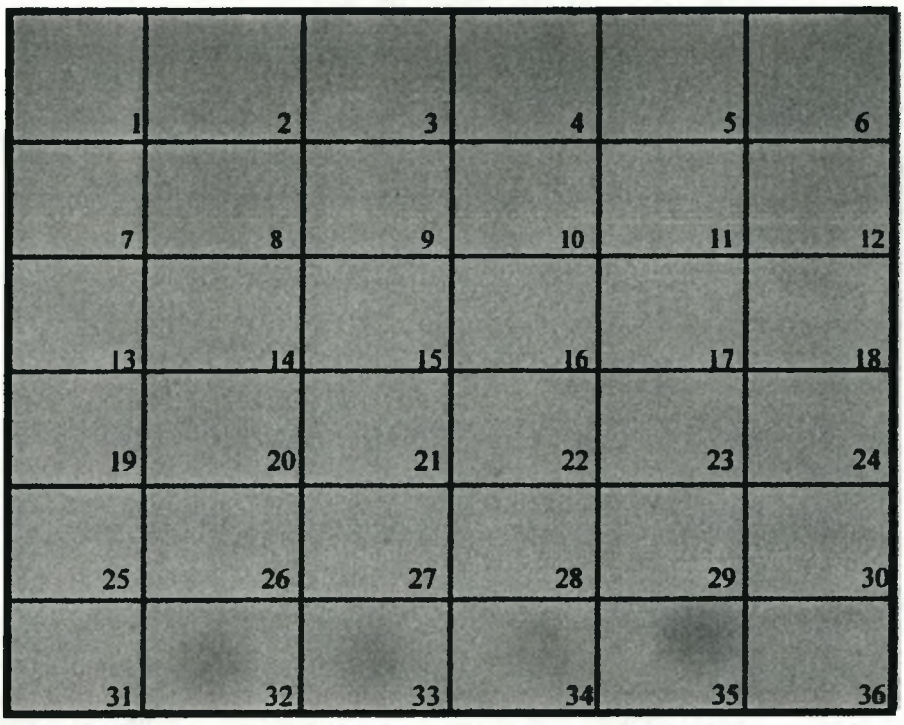


Figure 3.18 Autoradiograph of a dot blot probed radioactively with α-^{32}P[dCTP] G-protein cDNA probe.

Lanes:				
	1. C14187	11. C24680	21. C23078	31. water control
	2. C18073	12. C29025	22. C34759	32. **genomic DNA**
	3. C18618	13. C27268	23. C30460	33. **C19 RHG12**
	4. C222363	14. C8946	24. C32617	34. **C19 RHOrim-7**
	5. C15217	15. C15743	25. C32617	35. **Probe DNA**
	6. C22135	16. C20381	26. C24833	36. no sample
	7. C14690	17. C27993	27. C28410	
	8. C33205	18. C19161	28. no sample	
	9. C24493	19. C24596	29. no sample	
	10.C34439	20. C10501	30. no sample	

C, cosmid; C19RH, chromosome 19 radiation hybrid . Position of clones at the PFHBI locus see figs. 1.11 and 1.12. Positively-hybridising DNA samples are shown in **bold.**

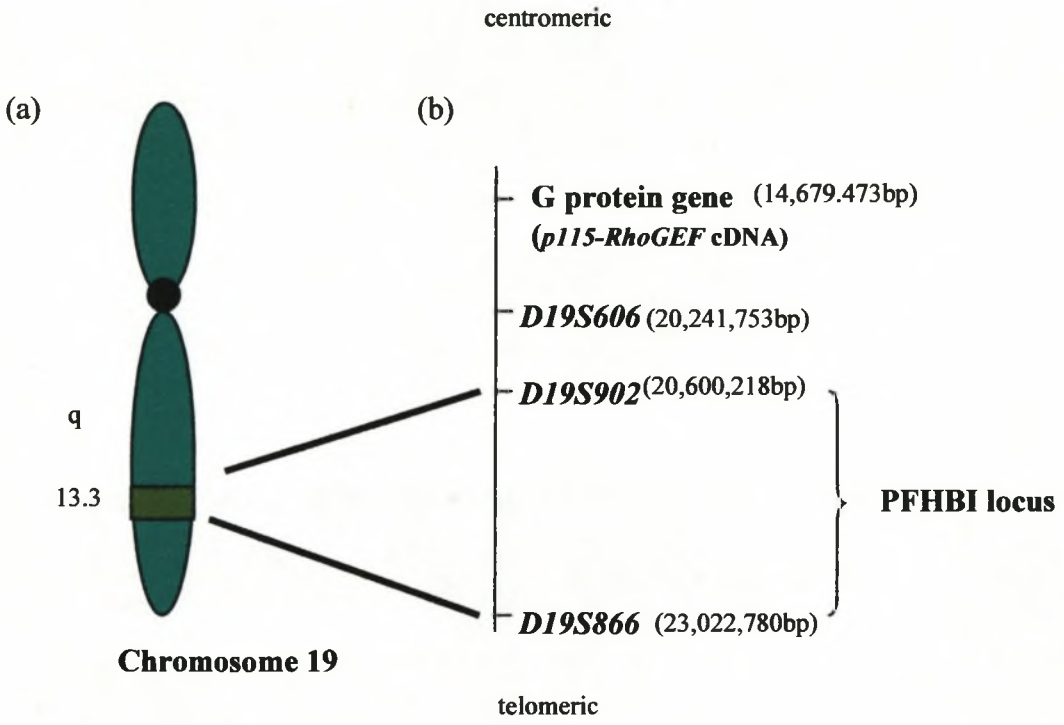


Figure 3.19 Schematic map showing the position of the G protein gene, *p115-RhoGEF* cDNA, relative to the PFHBI locus.

(a) Schematic drawing of chromosome 19 showing the location of the PFHBI locus at C19q13.3. (b) Integrated physical and genetic map showing the positions of the PFHBI locus and the G protein gene. Base pair positions (bp) of *p115-RhoGEF* cDNA and genetic markers *D19S606*, *D19S902* and *D19S866* were obtained from BLAST searches of the human genome database.

3.5.3 CTG repeat expansion

The results generated from employing the RED technique to analyse five PFHBI-affected samples, as described in section 2.14.1, performed in the Schalling laboratory, Karolinska Institute, Sweden, are shown in fig 3.20. Eleven controls were used, namely, three mildly affected DM patients, three DM patients and the genomic DNA from five unrelated healthy individuals (data not shown). A 120bp RED product is usually present in the general population and is considered normal. However, with sample 2 (2.III.9) larger products ranging from 150bp to 270bp and with sample 4 (2.IV.1) a 150bp product, were observed. The additional products were found to be due to expansion of a non-pathogenic polymorphic CTG repeat (AF020274) which maps to chromosome 18 (personal communication, Quiping [Schalling group]).

3.6 MUTATION SCREENING OF PFHBI CANDIDATES

The candidate genes selected for mutation screening, discussed in section 2.14, were *KCNA7*, *GSY1*, *BAX*, *KIR2.4* and *LIN-7B* (sections 3.6.1 and 3.6.2).

3.6.1 *KCNA7*

In this study, both, exons and the flanking intronic junctions of *KCNA7* were screened for disease-causing mutations (section 2.14.4), using the primer sets shown in table 2.3 (section 2.5).

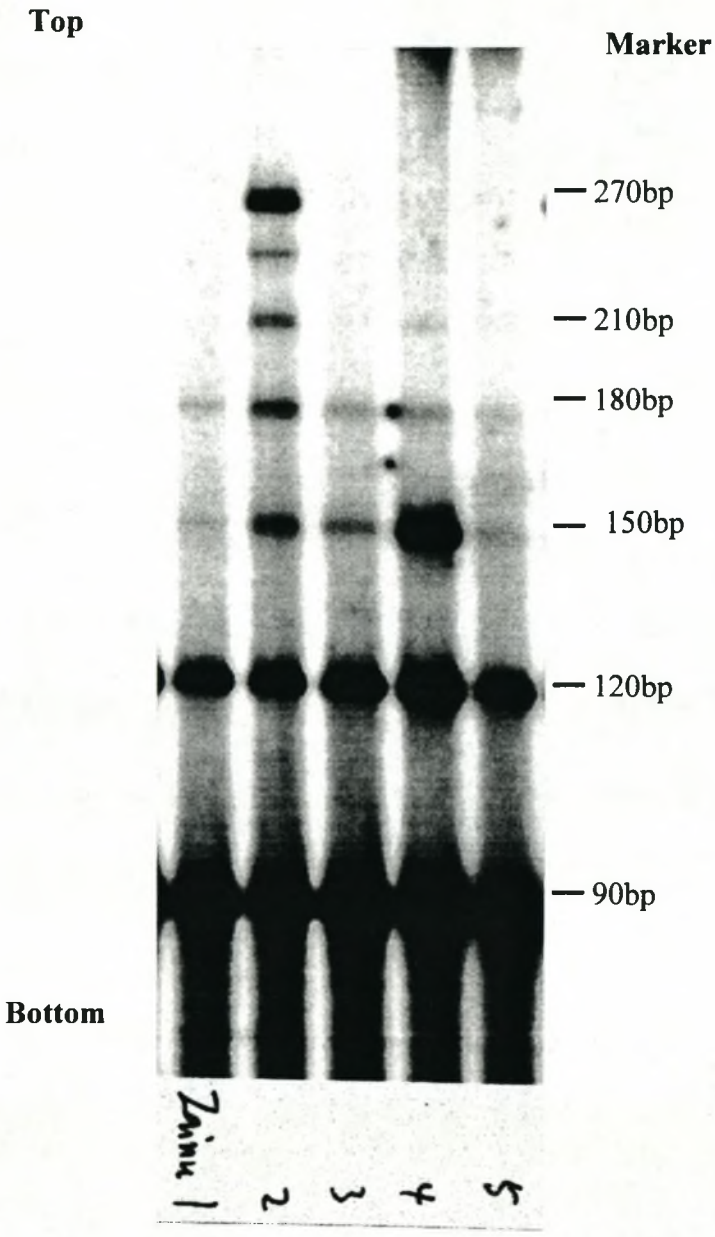


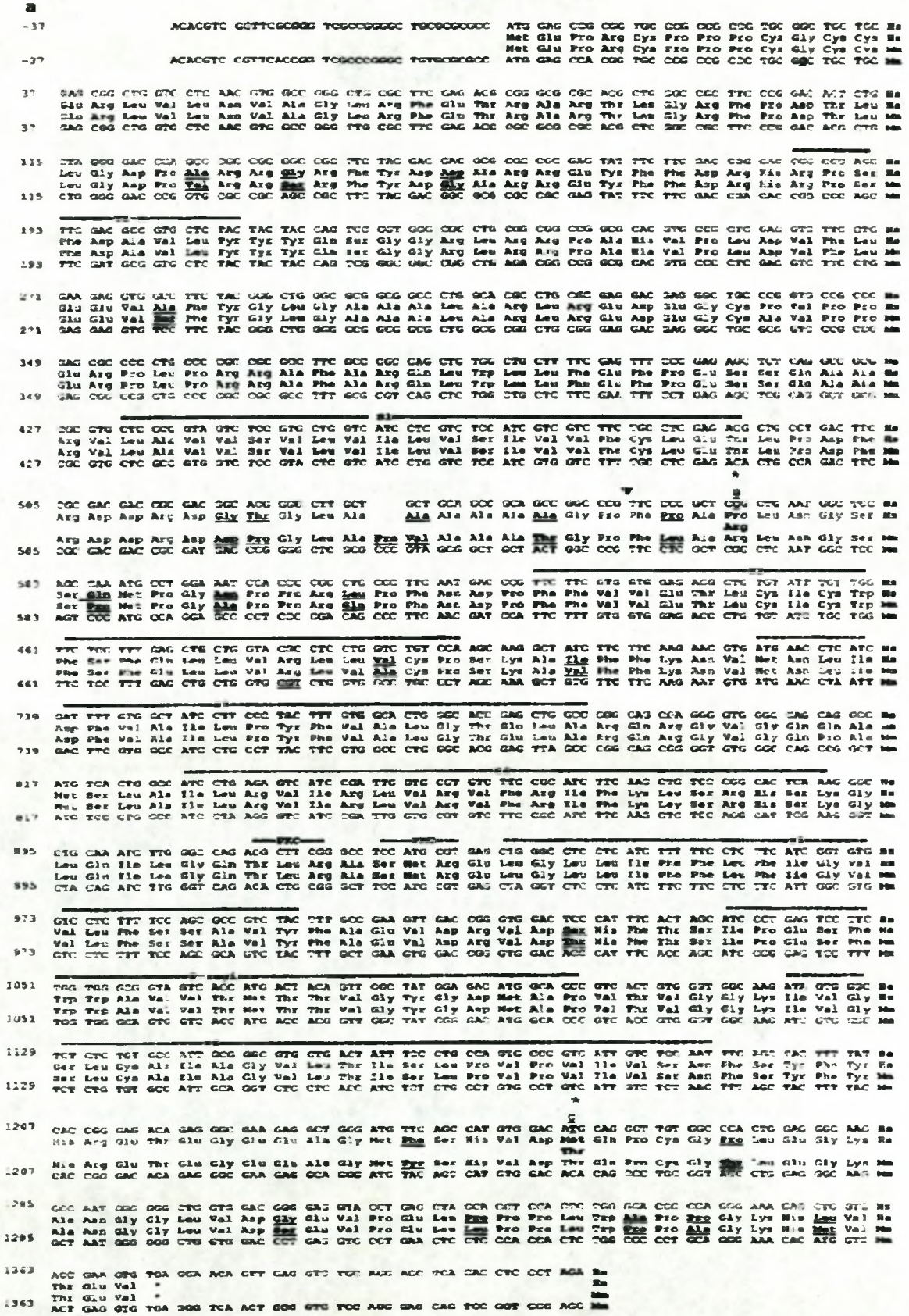
Figure 3.20 Autoradioagraph of DNA samples from PFHBI-affected individuals using the RED technique.
RED products, separated by gel electrophoresis, for each of the five PFHBI-affected individuals analysed were transferred to a membrane and probed with a labeled $(CTG)_{17}$ oligonucleotide. Lanes 1-5, DNA samples from PFHBI-affected individuals. Product sizes in bp are shown on the right of the autoradiograph. All the samples have a strong 120bp product. Sample 2 has five fainter products, ranging from 150bp to 270bp. Sample 4 has a strong 150bp product. Samples are from pedigree 2: 1, 13.II.1; 2, 2.III.9; 3, 4.IV.4; 4, 2.IV.1; 5, 4.IV.4 (personal communications, Quiping [Schalling group]).

3.6.1.1 Characterisation of the coding sequences and the genomic organisation

A BLAST search of the Genoscope database (http://www.genoscope.org) revealed that human genomic sequences (R0AA003ZF08G1) had homology to the mouse *kcna7* cDNA (AF032099). Another BLAST search against Genbank using the inferred human *KCNA7* sequence identified significant matches with human BAC clone BC52309 (AC0008687), mouse genomic contig (AC073711) and three placental expressed sequence tags (ESTs) AA021711, AI322534, AI324179) (Bardien-Kruger *et al.*, 2002). Using the program EST2GENOME (http://ftp.sanger.ac.uk/pub/EMBOSS), the coding sequence of *KCNA7* was deduced by aligning the mouse cDNA sequence (AF032099) with the human BAC clone BC52309 (AC0008687), retrieved from the databases, and the *KCNA7* sequence was deposited into the Genbank database (accession number AF315818). Concomitantly mouse *kcna7* sequence was corrected at several positions based on the alignments between published cDNA sequence (AF032099), the mouse genomic sequence (AC073711) and mouse EST sequences (fig. 3.21) (Bardien-Kruger *et al.*, 2002).

3.6.1.2 SSCP-PCR analysis

Two sets of overlapping primers spanning exon 1 and five sets spanning the larger exon 2 (table 2.3 and fig. 2.13) were used to screen the coding regions of *KCNA7* for pathogenic mutations using both SSCP-PCR analysis and direct sequencing, as described in sections 2.6.3 and 2.7.2, respectively. The SSCP-PCR results for exons 1 and 2 are shown in



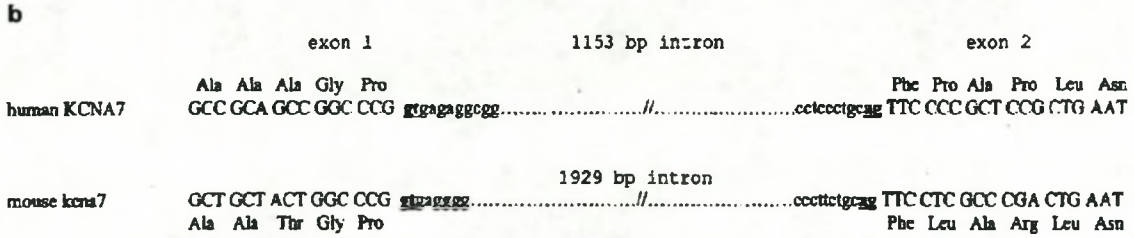
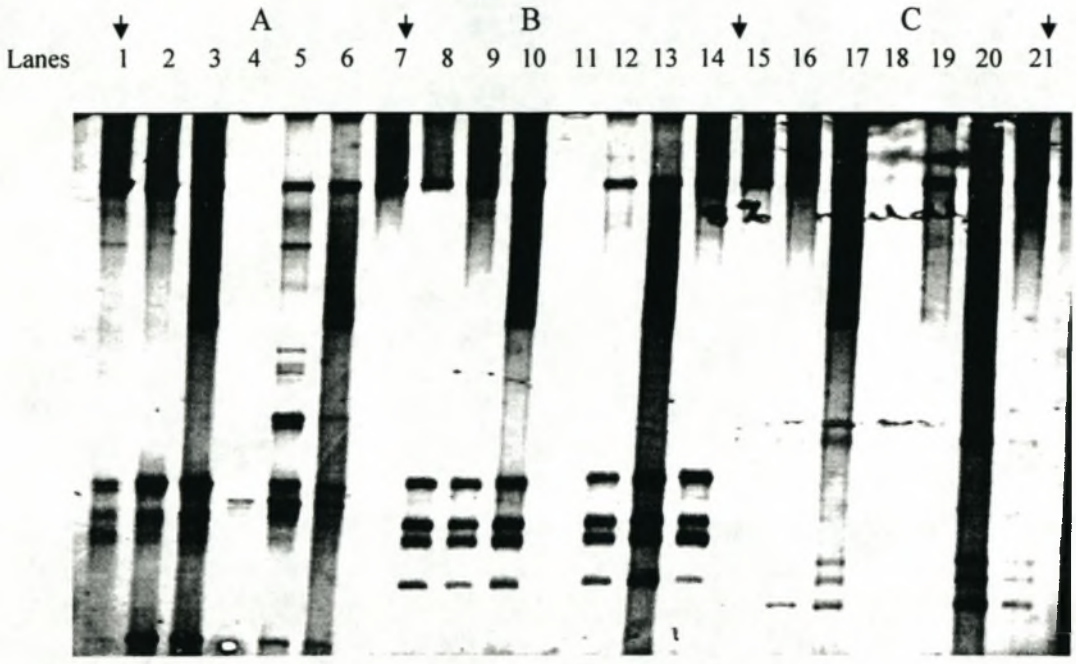


Figure 3.21 Comparison of human *KCNA7* with mouse *kcna7*.
(a) Comparison of the nucleotide and deduced amino acid sequences of human KCNA7 and mouse kcna7. The six membrane-spanning domains (S1 through S6) and the pore-forming regions (P-region) are indicated. The arrowhead (▼) indicates the location of the single intron that lies between S1 and S2. Differences between the human and mouse amino acids are in bold and underlined. The two polymorphisms in the human gene are indicated (*) at nucleotide position 566 C:G and 1253 T:C of *KCNA7*. The mouse *kcna7* sequence has been corrected at several postions based on the alignment between the published cDNA sequence (AF032099), the mouse genomic sequence (AC073711) and mouse EST sequences. The missing 'G' at position 29 in AF032099 is indicated in bold and underlined.
(b) Exon-intron junctions of human *KCNA7* and mouse *kcna7*. The human 1153bp intron and the mouse 1929bp intron are shown (Bardien-Kruger *et al.*, 2002).

figs 3.22 and 3.23, respectively. No disease-causing mobility shifts were found between affected (1.V.10, 6.IV.12, 9.IV.5 and 5.III.1) and unaffected PFHBI (1.V.6 and 9.III.6) family members. However, two changes in banding pattern were observed in an unaffected PFHBI individual (9.III.6) with primer sets D and G. This was further investigated in kindred 9 of pedigree 2.

3.6.1.3 Sequencing data

The PCR products A, B, C, D, E, F and G of affected (1.V.10, 9.IV.5 and 5.III.1) and unaffected PFHBI (1.V.6 and 9.III.6) individuals were sequenced using the ABI310 system. The DNA sequencing data did not reveal any pathogenic mutations.

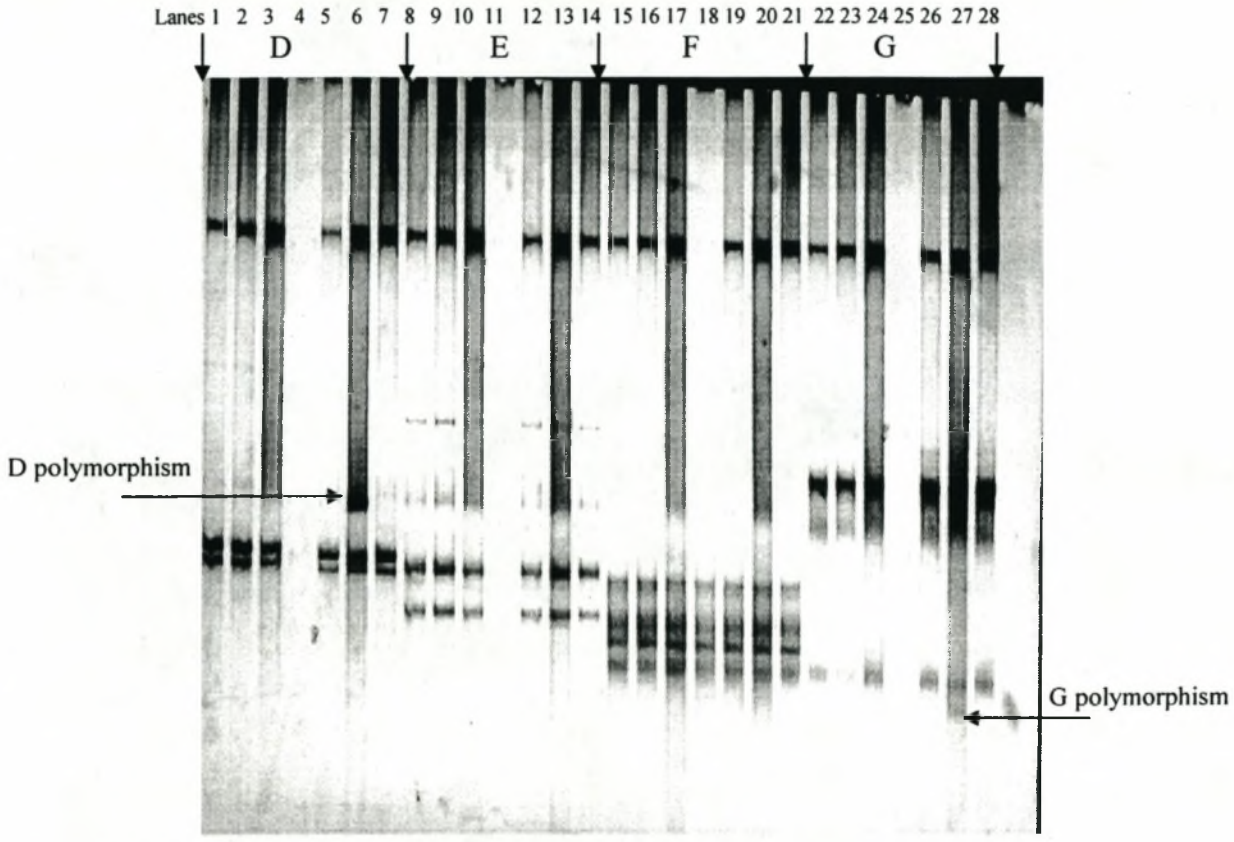


Denatured samples of individuals:

1. 1.V.10 (A)
2. 6.IV.12 (A)
3. 9.IV.5 (A)
4. 5.III.1 (A)
5. 1.V.6 (U)
6. 9.III.6 (U)
7. 13.II.8 (S)

Figure 3.22 PCR-SSCP mutation screening of exon 1 and 5'-end of exon 2 of *KCNA7*.
PCR products A and B were obtained by PCR amplification of exon 1 using primer sets A and B (table 2) and DNA from individuals 1 to 7. PCR products C were obtained by amplification of 5'-end of exon 2 using primer set C (table 2.3) and DNA from individuals 1 to 7. PCR-SSCP samples were run on a 8% polyacrylamide with glycerol: Lanes 1 to 7 contain heat denatured samples from part A of exon 1, lanes 8 to 14 contain heat-denatured samples from part B of exon 1 and lanes 15 to 21 contain heat-denatured samples from part C of the 5'-end of exon 2. A = PFHBI-affected. U = unaffected first-degree relative. S = unrelated spouse.

Exon 2



Denatured samples of individuals:

1. 1.V.10 (A)
2. 6.IV.12 (A)
3. 9.IV.5 (A)
4. 5.III.1 (A)
5. 1.V.6 (U)
6. 9.III.6 (U)
7. 13.II.8 (S)

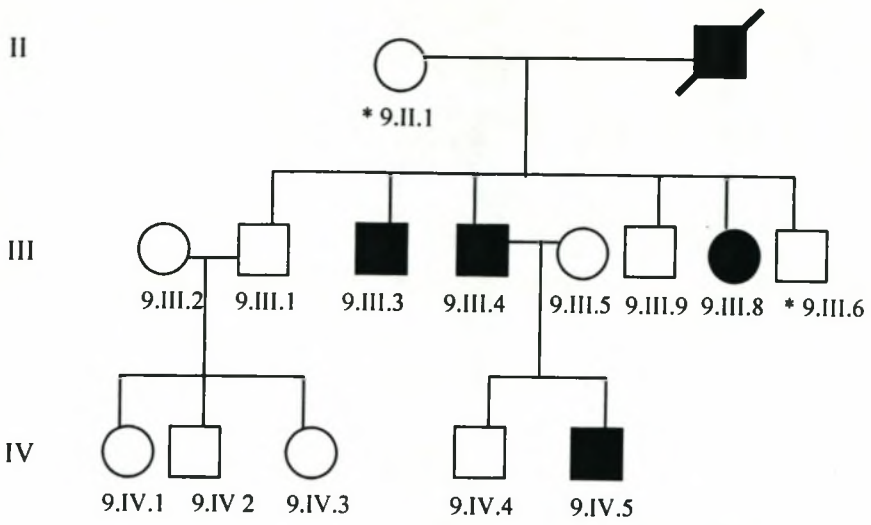
Fig 3.23 PCR-SSCP mutation screening of exon 2 of *KCNA7*.
PCR products were obtained by PCR amplification of exon 2 using primer sets D, E, F and G (table 2.3) and DNA from individuals 1 to 7. PCR-SSCP samples were run on a 8% polyacrylamide with glycerol. Lanes 1 to 7 contain the heat-denatured samples of part D of exon 2, lanes 8 to 14 contain the heat-denatured products of part E of exon 2, lanes 15 to 21 contain the heat-denatured products of part F of exon 2 and lanes 22 to 28 contain the heat denatured samples of part G of exon 2. The bold arrows on the left and the right show the single strand conformers for each part of exon 2. The fine arrows indicates the D polymorphism in sample 6 and the G polymorphism in sample 27. A = PFHBI-affected. U = unaffected first-degree relative. S = unrelated spouse.

3.6.1.4 Polymorphism analysis

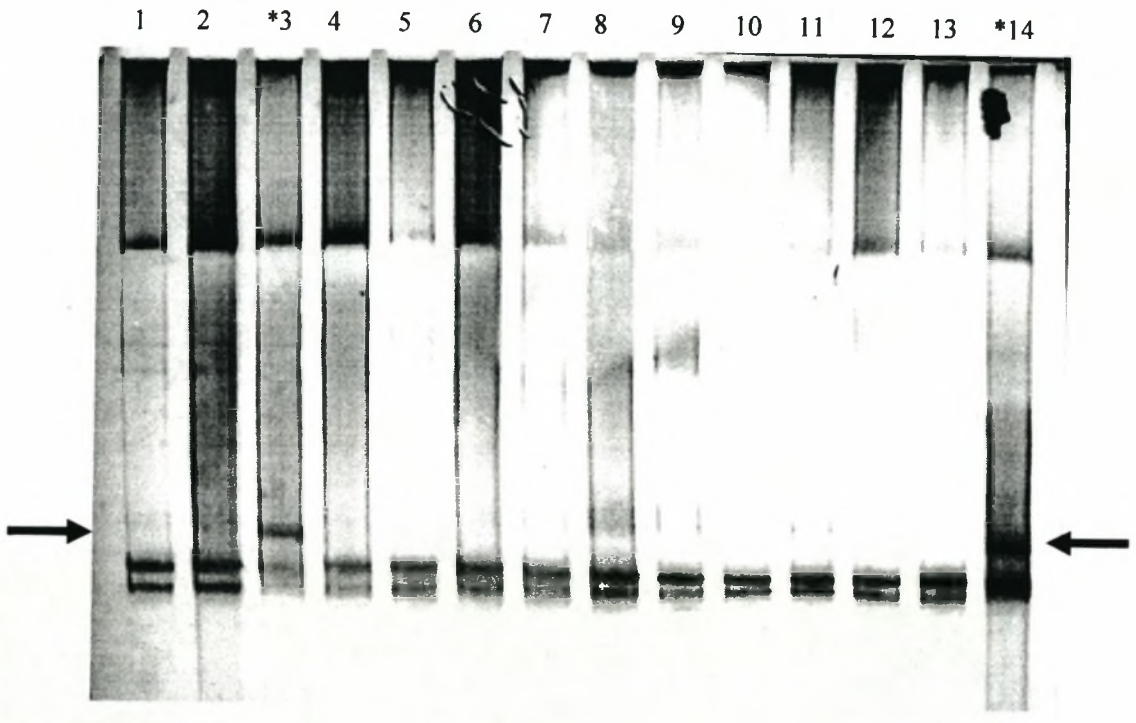
Primer sets D and G were used to screen all the members of kindred 9 of pedigree 2. The PCR-SSCP results using primer set D are shown in fig. 3.24. In both cases, using primer sets D and G, mobility shifts were observed in only two unaffected individuals (9.II.1 and 9.III.1), namely, "a married in" and her son. The PCR products of both individuals were sequenced and are shown in fig. 3.25. In case of the D polymorphism, a C to G nucleotide change was observed at position 566, which caused an amino acid change from proline to arginine (fig. 3.25a). In case of the G polymorphism, a T to C nucleotide change was observed at position 1253, but no amino change occurred (fig. 3.25b). The allelic frequencies of the two exon 2 SNPs determined in a general South African Afrikaner population were 0.62:0.38 for nucleotide 566 C:G polymorphism, respectively, and 0.7:0.3 for the nucleotide 1253 T:C polymorphism, respectively. The two novel SNPs were submitted to the SNP database, NCBI and were designated AF315818-566 and AF315818-1253 (table 3.14).

3.6.2 *KIR2.4, LIN-7B, BAX and GSY1*

A summary of the mutation screening results generated for the exons and flanking intronic junctions of *KIR2.4, LIN-7B*, *BAX* and *GSY1* are presented in tables 3.14 and 3.15. The details of the exonic organisation of the candidate genes are shown in section 2.14 and the clustal alignment results of each exon for each gene are shown in appendix III. A previously reported SNP in exon 7 (rs5464) (Genbank Acc no. Z33610) of *GSY1* was found in the DNA of an unaffected and an affected individual from pedigree 1



Kindred 9 of pedigree 2



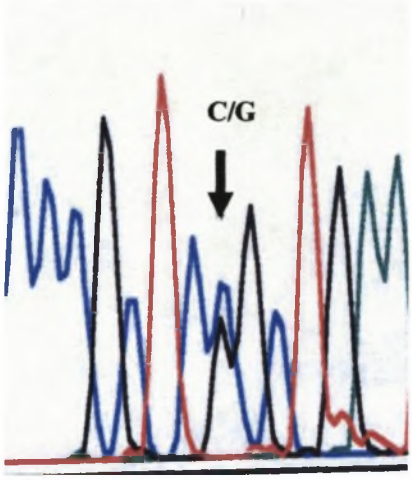
Lanes:

1 = 9.III.3
2 = 9.III.1
***3 = 9.II.1**
4 = 9.IV..1
5 = 9.IV.2
6 = 9.IV..3
7 = 9.III.3
8 = 9.III.4
9 = 9.III.5
10 = 9.IV.4
11 = 9.IV.5
12 = 9.III.9
13 = 9.III.8
***14 = 9.III..6**

Clinically unaffected
Clinically affected (A)
Deceased affected

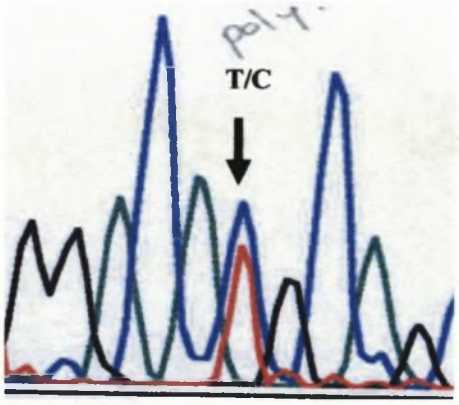
Figure 3.24 PCR-SSCP analysis of kindred 9 of pedigree 2 with primer set D.
The lane numbering (1-14) on the gel corresponds to heat-denatured PCR amplified products using primer set D (table 2.3, fig. 2.13). The PCR-SSCP producects were run on 8% polyacrylamide gel. *, Mobility shifts are visible in samples 3 (9.II.1) and 14 (9.III.1). The position of the shifts are indicated by the arrows.

505 CCC GCT CCG CTG AAT GGC TCC
Pro Ala Pro Leu Asn Gly Ser
CCC GCT CGG CTG AAT GGC TCC
Pro Ala Arg Leu Asn Gly Ser



(a) Polymorphism in D

1207 AGC CAT GTG GAC ATG CAG CCT
Ser His Val Asp Met Gln Pro
AGC CAT GTG GAC ACG CAG CCT
Ser His Val Asp Met Gln Pro



(b) Polymorphism in G

Figure 3.25 Sequencing traces showing the D and G polymorphisms of exon 2 of *KCNA7*.
(a) A trace showing the polymorphism with primer set D, a nucleotide change of C to G. (b) A trace showing the polymorphism using primer set G, a nucleotide change of T to C. The changes are shown in **bold** and the polymorphism is indicated by a vertical arrow. The nucleotide sequence with the corresponding amino acids are also shown. The afffected amino acid is shown in red in both cases.

(table 24 in Appendix III). The alleles observed were C/T in nucleotide position 1186 of the Genbank nucleotide sequence accession number J04501. In addition, a number of errors were found in previously submitted sequence data. A previously identified SNP (rs1057369) was found in the submitted database sequence of exon 6 (L22473) of *BAX* at base pair position 552 (tables 3.14 and 3.15). The BLAST search results of this sequence against the human genome database indicated that the G at base pair position 1756471 on the NT_011109 contig should be replaced by an A (table 16 in Appendix III). This change may very well represent a SNP and the allelic frequency in different unaffected ethnic populations should have been determined. However, no disease–causing mutations were found in any of the exons of *BAX* screened.

Table 3.14 Summary of mutation screening data

Gene	Exon No	SNP/Polymorphisms/ errors detected	Reference
KCNA7			
	1	none	fig. 3.22
	2	2 submitted SNPs, AF315818-566 and AF315818-1253	figs. 3.23 and 3.25
KIR2.4			
	1	none	table 1
	2	none	tables 2 & 3
	3	none	tables 4, 5 & 6
LIN-7B			
	1	none	table 7
	2	none	table 8
	3	none	table 9

table 3.14 cont

	4	none	table 10
	5	none	table 11
BAX			
	1	none	table 12
	2	none	table 13
	3	none	table 13
	4	none	table 14
	5	none	table 15
	6	SNP in exon6 (rs1057369) (already described)	table 16
GSYI			
	UTR5A	an error in the submitted sequence, S79190, CC	table 17
	UTR5B	none	table 18
	UTR5C	none	table 19
	1	none	table 20
	2	none	table 21
	3	none	table 22
	6	none	table 23
	7	SNP in exon 7 (rs5464) (already described)	table 24
	8	none	table 25
	9	none	table 26
	10	none	table 27
	13	none	table 28
	14	none	table 29
	15	none	table 30
	16	none	tables 31, 32 and 33

Tables 1 to 33 are found in appendix III.

Table 3.15 Summary and comparison of submitted sequence variations identified in the present study and other studies.

Candidate gene	SNP/ error	Amino acid substitution	Allelic frequency South African Afrikaaner population	Present study (Acc No.) date of submission	Other NCBI submissions date of submission
KCNA7	C>G SNP	Pro566Arg	0.620.38	Brincor | KCNA7-5666 ss2419735 31/10/00	Yusuke IM | IMS JST16932 ss4996592 12/8/02 CSHL-HAPMAP | CSHL-HuAA-200402 ss16790996 17/2/04 SSAHASNP | WGSA-200403 ss21550400 20/3/04
	T>C SNP	synonomous	0.7:0.3	Brincor | *KCNA7* 1253 31/10/00	TSC-CSHL | TSC0220257 ss1481085 ss2419736 7/9/00 Yusuke | IMS-JST169930 ss49965920 12/8/02 WI_SSAHASNP ss6628714 12/2/03 CSHL-HAPMAP ss17613181 19/2/04 SSAHASNP ss21521607 20/3/04
BAX	G>A	synonomous	not determined	-not submitted	 LEE | 718434 ss1537999 13/9/00
GSY1	CC>GG error	UTR5A in S79190		-not submitted	
	C>T SNP	exon 7 1186 in J04501	not determined	not submitted	rs5464 Aravinda/00543 ss6929 29/6/99 WIAF-CSNP/WIAF-12181 Ss3173163 19/6/01

Acc, accession number; C, cytosine; G, guanine; T, thymine; Pro, proline; Arg, arginine; *KCNA7*, potassium voltage-gated channel shaker-related subfamily member 7; *BAX*, *BCL2*-associated X protein; *GSY1*, glycogen synthase 1

The database sequence of UTR5A of *GSY1* (acc no. S79190) (http://www.ncbi.nlm.nih.gov/nucleotide) at positions 139 and 140 were incorrect (table 3.14). It was found that the nucleotides CC at these two positions should be replaced with nucleotides GG (table 17 in Appendix III). This result was verified by a BLAST homology search of the UTR5A database sequence against the Human genome database (http://www.ncbi.nlm.nih.gov/human genome). It is also possible that the sequence changes may represent SNPs and the allelic frequencies should have been determined in different ethnic population groups.

From the alignment results of *KIR2.4, LIN-7B*, *BAX* and *GSY1*, no sequence changes were found when comparing the coding sequences, and the exon/intron boundary sequences of affected to unaffected DNA samples from the PFHBI panel therefore, these regions were excluded as disease-causing.

CHAPTER 4

DISCUSSION

INDEX

4.1 THE STUDY DESIGN

Previous studies in the search for the PFHBI gene involved the use of linkage analysis and genetic fine mapping, resulting in the placement of the PFHBI locus on chromosome 19q13.3 in a 10cM interval (Brink *et al.*, 1994; Brink *et al.*, 1995). Subsequent fine mapping further reduced the PFHBI locus to a 7cM interval (fig. 3.1) (personal communication, De Jager, 1998).

In this study, a multi-step strategy was used to continue the search for the PFHBI disease-causing gene, which involved further fine mapping and reduction of the PFHBI locus using published and novel polymorphic markers (section 3.1). In the early stages of this project (June 1999), very few annotated genes had been deposited into the publicly available databases; therefore, EST databases were scanned with a view to identifying as many expressed genes at the PFHBI locus as possible (section 3.2). An integrated physical map of the PFHBI locus was generated in order to facilitate the search for the PFHBI gene and the map was updated periodically, as new information was deposited into the databases (section 3.3).

Affected members of the PFHBI family were screened for the presence of any unusually expanded triplet repeat (CTG)-containing genes, while the PFHBI locus was screened for the presence of a G protein-encoding gene and a *Cx* gene, as they were considered plausible functional PFHBI candidate genes (section 3.5). As the HGP progressed (May 5, 2003), many more genes were identified and deposited into the databases and, as part

of the present study, genes at the PFHBI locus were characterised and prioritised as PFHBI candidates. A priority list for the mutation screening of PFHBI candidate genes was drawn up (section 3.4) and selected genes were screened for the disease-causative mutation, as part of the present study, in a panel of selected subjects (section 3.6).

The panel of PFHBI-affected and unaffected family members was selected based on strict criteria with unequivocal clinical features as described in section 2.2. In addition, the clinically affected individuals chosen carried the disease-associated haplotype, whereas the controls were unaffected first degree relatives and "married-in" individuals, who had a normal ECG and did not carry the disease-associated haplotype. Although sequence variants occurred in the candidate genes screened in this study, they were present in PFHBI-affected individuals, in unaffected family members and in the general population (tables 3.14 and 3.15); thus no disease-causing mutation was found.

During the course of the study, the associated region was reduced to 4cM. In addition, novel genetic markers and single SNPs were deposited in databases (NCBI). All the BLAST searches used in the generation of the integrated physical map were performed against the final chromosome 19 contig NT_011109. This contig sequence data, which existed in May 2003, forms part of the final chromosome 19 sequence published in April 2004 by Grimwood *et al.*

4.2 REDUCTION OF THE PFHBI LOCUS

Finding a disease-causing gene in a 4cM target region is a challenging task, success more often follows when the region has been refined to a more manageable distance of approximately 1-2cM. For example, RP13- and the Pendred syndrome-associated genes were identified when the relevant loci were reduced from 3 to 1cM (McKie *et al.*, 2001) and 2.5 to 1.7cM (Coucke *et al.*, 1997), respectively.

Furthermore, limiting the size of the locus was especially relevant, as chromosome 19 is particularly gene-rich and harbours an average of 26 genes per Mb (Grimwood *et al.*, 2004), compared, for example, to 6.5 genes per Mb on chromosome 13 (Dunham *et al.*, 2004). During the present study, the estimated number of genes at the PFHBI locus was reduced by 34%, i.e., from a predicted 96 to a predicted 63 genes.

Further reduction of the region requires the investigation of more family members, or new families of either South African origin or from other countries. During the present study, new family members of pedigree 2 were genotyped but they did not yield any new recombination events. In addition, a collaboration was established with a French group working on a Lebanese family with a conduction disease designated CCD (De Meeus *et al.*, 1995), which was subsequently considered to be the same disease as PFHBI (personal communication, Brink, 2000) (section 1.3.2.3.4). An integrated map generated from the different markers used by the two groups (Brink *et al.*, 1995; De Meeus *et al.*, 1995) showed that the two intervals overlapped (fig. 1.9) (personal communication, Corfield,

2000). Using genetic fine mapping, the PFHBI locus was reduced to 7cM between markers *D19S412* and *D19S866* (personal communication, De Jager, 1998). Unfortunately, the collaboration between the French and Brink/Corfield groups has lapsed.

Tetranucleotide markers

At the start of the present study, other PFHBI team members were developing dinucleotide markers from cosmid clones spanning the PFHBI locus, whereas part of this study focused on the development of tetranucleotide STRs from cosmid clones. The tetranucleotide STRs reported to occur most abundantly in the human genome are A_3G (38%), AGAT (29%), AAGG (10%) and AAAT (5%) (Utah Marker Development Group, 1995). Based on these findings, $(A_3G)_n$ STRs were selected for marker development as part of this study. Since the sequence of the PFHBI region is now readily available, a more feasible approach would be screen the region for commonly occurring repeat motifs for marker development using a bioinformatics tool such as Repeatfinder (Benson, 1999) or to scan for new polymorphic markers which may have recently been deposited into the public database.

Selection of the cosmids for $(A_3G)_n$ STR development was based on their position relative to the genetic map of the locus (fig. 3.3). It was hoped that the ones selected were close to recombination breakpoints detected during fine mapping and thus might serve to refine the search interval. On sequencing, it was found that cosmid insert 29395

contained an $(A_3G)_n$ tetranucleotide repeat, however, further development of the marker was problematic because the repeat motif was embedded in a region containing *Alu* repeat sequences (figs. 3.6 and 3.7). This finding was not unexpected, since it is reported that *Alu* repeats occur at an average frequency of one copy per 4kb, in genomic sequence (Strachan and Read, 2000). It has also recently been reported that for chromosome 19 *Alu* repeats make up 25.8% of the total sequence, compared to 13.8%, 13.3%, 9.5% and 16.8% on chromosomes 7, 14, 21, and 22, respectively (Grimwood *et al.*, 2004). The disadvantage of designing a primer from within an *Alu* repeat sequence is the generation of a high background due to non-specific priming from the other similar sequences within the human genome (Beckman and Weber, 1992). For this reason, the marker within cosmid insert 29395 was not developed further.

Nucleotide sequencing of cosmid insert 18618, which had been given a positive signal on hybridisation to an $(A_3G)_n$ probe, failed to reveal any $(A_3G)_n$ repeat motifs. The probable explanation is that the motif was lying within the un-sequenced region of the insert, since only about 500bp, adjacent to the forward and reverse Bluescript primers, of the 1.7kb insert (fig. 3.4) was sequenced. Assuming this to be the case, finding the $(A_3G)_n$ repeat motif would have involved the design of new primers from the sequences obtained with the vector-specific primers. However, the development of this marker was not pursued at this time as other strategies were being followed by the PFHBI team. At a later stage, when draft sequence became available, a bioinformatics approach was used and

dinucleotide repeat sequences from BAC and cosmid clones spanning the PFHBI locus were developed as $(CA)_n$ markers (section 3.1.5).

Dinucleotide markers

In the present study, the draft sequences of 27 BAC and cosmid clones spanning the PFHBI locus were searched for the presence of $(CA)_n$ repeat motifs using Tandem Repeatfinder (fig. 3.10) (Benson, 1999). The strategy was to develop markers from BAC and cosmid clones centromeric to *D19S596*, and telomeric to *D19S604* (fig. 3.10), in order to refine the PFHBI locus. Four markers were successfully developed, and showed a Mendelian pattern of inheritance (fig. 3.11). The deposited database sequences from which markers *G68141*, *G68142*, *G68143* and *G68144* were developed contained runs of CA repeat motifs of 20, 12, 11 and 6, respectively, which, in turn, when genotyped in a general population, generated 16, 4, 2 and 1 alleles, respectively (Tables 3.2, 3.3 and 3.4). This was in agreement with the findings of Weber (1990) that the length of the $(CA)_n$ repeat motif is an indicator of the extent of polymorphism (see section 1.4.1.1).

Genotyping with *G68141*, *G68142* and *G68143*

Genotyping with two of the novel markers developed, *G68141* and *G68142,* in PFHBI-affected individuals in kindred 4 of pedigree 2 (section 3.1.5, fig. 3.12) showed that the centromeric breakpoint of the disease-associated haplotype lies between *D19S902* and *G68141*. The disease-associated haplotype (*G68141-G68142-D19S596-D19S604-D19S866)* is thus maintained at 4cM or at the equivalent physical distance of 2.42Mb

(section 3.1.4 and fig. 3.8). In order to precisely delineate the recombination breakpoint, a search for STR markers lying between *D19S902* and *G68141* could be undertaken. This region represents a physical distance of only 447kb, but refining the interval could exclude up to five genes in this gene-rich location, namely, *CABP5*, *PLA2G4C*, *LIG1*, *CARD8* and *EMP3* (fig. 3.15), depending on exactly where the breakpoint lies.

As part of another study by a member of the PFHBI team, a recombination breakpoint occurring at the telomeric end of the PFHBI locus in kindred 13 of pedigree 2, i.e., between *D19S604* and *D19S866* (fig 3.1), was investigated (Du Plessis, 2000) (data not shown). The marker was used, *G68143*, in an attempt to narrow the locus was typed in the individuals from kindred 9 and 13 of pedigree 2, the former kindred serving as the "control". However, no allelic differences were found between individuals of kindred 9 and 13 of pedigree 2, suggesting that the breakpoint was nearer the telomeric limit.

Identification of new recombination events

As mentioned earlier, the acquisition of new PFHBI family members in pedigrees 1, 2 and 5 may identify new recombination events, which may lead to further reduction of the disease region, thereby reducing the number of PFHBI candidate genes to be screened. Since the genetic distance of the PFHBI locus is 4cM, the possibility of finding more recombination events is possible. It is, therefore, important to continue to trace new PFHBI individuals from pedigrees 1, 2 and 5 and to genotype them across the critical interval. For this purpose, a relationship between the research clinician, the genetic sister

(Sr Althea Goosen) and the PFHBI families has been maintained to keep track of new births. It would also be useful to re-establish links with the French (CCD) group, or any group from other countries that have mapped inherited cardiac conduction diseases to chromosome 19q13.3.

4.3 SEQUENCING OF CHROMOSOME 19

The sequencing of chromosome 19 lagged in comparison with the sequencing of the other chromosomes (IHGSC, 2001). The first physical map of chromosome 19 was published in 1995 (Ashworth *et al.*, 1995) and, at that stage, sequencing of chromosome 19 was in the initial stages. In February 2002, 50% of the region containing the PFHBI locus was still in draft form and was composed of three smaller contigs, NT_011190, NT_011140 and NT_011157 (fig. 3.13). In 2002, the Stanford Human Genome Center took over the sequencing of chromosome 19 and in April 2004, 99.9% of chromosome 19 was sequenced and the analysis of the most recent map was published (Grimwood *et al.*, 2004). By then, the nucleotide sequence data across the PFHBI locus had been merged into one contig designated NT_011109 (May 2003) (fig. 3.15).

Limitations of draft sequence

While developing STRs from draft sequence, a number of discrepancies in both the reported coverage and the accuracy of mapping of genomic clones were found, this suggested some limitations in the use of the draft genome sequence to provide accurate

positional information and detail of chromosomal regions. For example, in the present study, it was found from BLAST searches of a 500bp sequence of clone AC010619 that it mapped to two places on contig NT_011109, namely, within the PFHBI locus, as well as outside the telomeric limit of the PFHBI locus (see section 3.1.5) (fig. 4.1). Katsanis and colleagues (2001) reported similar problems when they evaluated the quality of the draft human genome sequence of a representative region using experimental and computational methods.

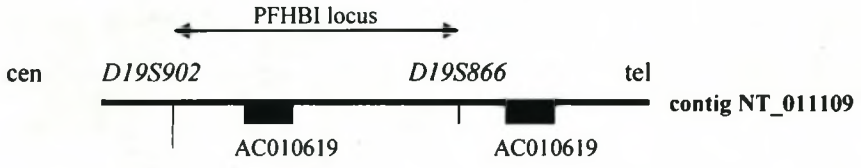


Figure 4.1 Positions of BAC clone sequences mapping within and outside the telomeric end of the PFHBI locus.
Black line shows contig NT_011109. Solid black box show the position of BAC clone AC010619 sequences on the contig NT_011109. Centromeric genetic marker *D19S902* and telomeric genetic marker *D19S866* delineate the PFHBI locus. cen, centromeric end; tel, telomeric end.

This was possibly due to a duplication of this portion of the AC010619 clone sequence on the chromosome, which would explain why the same sequence would map to two different positions on contig NT_011109 (IHGSC, 2001) (see section 3.1.5). This notion is supported by the finding that chromosome 19 sequence shows duplication, with 7.53% of the sequence sharing sequence homology with more than one location in the genome (Grimwood *et al.*, 2004). Another problem experienced during this study was that

primers generated from the insert of BAC clone AC008888 draft sequence for potential $(CA)_n$ STRs development, when tested experimentally using PCR amplification, did not generate the expected product but did amplify genomic DNA. One possible explanation is that LLNL had assigned the incorrect sequence to the clone, or that a clone switch had occurred (section 3.1.5). Consequently, the draft sequence of this clone was not used further for marker development.

The four successfully developed $(CA)_n$ markers were submitted to Genbank (see table 3.1). These novel markers may serve as a useful resource for other researchers searching for disease-causing genes in the neighborhood of the PFHBI locus, such as LGMD2 (Driss *et al.*, 2000) and *DFNA4* (Chen *et al.*, 1995), since their loci encompass these markers (section 1.4.4 and figs. 1.14 and 3.15). To date, the responsible disease-causative genes have not been identified.

4.4 GENE TRANSCRIPT SEARCHES

In the early stages of the project (June 1999), consensus sequences derived from the 38 STACK-clustered ESTs mapped to the PFHBI locus corresponded to only six proteins. Furthermore, not enough information regarding gene structure and tissue expression profiles of the corresponding genes was available and the few protein hits obtained at this

stage of the study gave an indication of the paucity of annotated proteins in the protein databases (section 3.2, table 3.5).

By May 2003, the GENEMAP99 and the LLNL site had still not updated their EST database, probably because in 2002 Stanford University had taken over the sequencing of chromosome 19. However, the same 38 ESTs identified in mid-1999 now yielded 23 protein matches, which is an indication of the increase in annotation of the proteins in protein databases. Of the 23 proteins, seven (*CA11*, *ZNF473*, *KPTN*, *FLJ25129*, *CRD8/TUCAN*, *EHD2* and *ATF5*) were listed as being expressed in heart.

Transcription factor mapping to the PFHBI locus

Application of gene prediction programmes showed that *ZNF473* has a transcriptional factor signature domain and on this basis is a plausible PFHBI candidate gene. A number of transcription factors, namely, MEF-2 and GATA-4, have been implicated in cardiac development (Franco *et al.*, 1998; Moorman *et al.*, 1998) (section 1.2.3) and NK2-5 and TBX5 have been associated with the cardiac diseases, HOS (Basson *et al.*, 1999) and ASD (Schott *et al.*, 1998), respectively. The position of *ZNF473* changed several times on the physical map during the course of this study and homology searches would sometimes place this gene close to the telomeric limit of the PFHBI locus (June 2000) and at other times outside the PFHBI locus (fig. 3.13) (February 10, 2002). In April 2004, a BLAST search showed that *ZNF473* maps across the most telomeric marker *D19S866* of the PFHBI locus (fig. 4.2). Unfortunately, insufficient time was available to screen

ZNF473, but it should be investigated at a later stage, specifically the part of the gene which lies within the PFHBI locus.

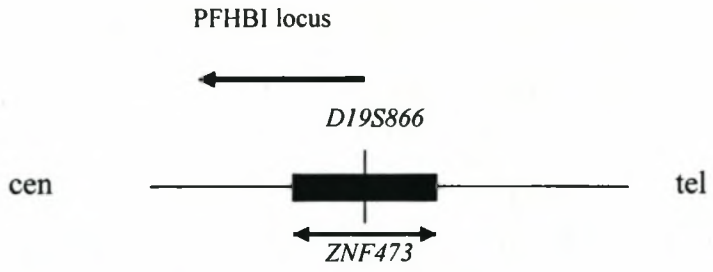


Figure 4.2 Position of the *ZNF473* relative to genetic marker *D19S866* (May 2004).
cen, centromeric; tel, telomeric. The PFHBI locus lies centromeric to *D19S866*.

Earlier gene searches using ESTs

In the earlier part of this study, plausible PFHBI candidate genes were identified using EST searches as discussed in sections 4.4 and 4.6 respectively. The retrieval of ESTs from Genemap99 and LLNL was necessary at the start of this study, since the dbEST database was in its infancy and very few genes had been placed at the PFHBI locus. In the later stages of the study (May 5, 2003), finished sequence covering the PFHBI locus, namely, contig NT_011109, became available and the updated sequence was used to generate an integrated physical and genetic map. The EST approach was a reasonable strategy during the earlier part of the study. If more sequence data had been available at that stage an alternative approach would have been to download this sequence and to perform a BLAST homology search of this data against the dbEST database. This would

have resulted in obtaining most of the then available ESTs mapping to the PFHBI locus. The next steps would have been to use clustering programmes (such as STACK and UNIGENE) and protein homology searches to find as many of the encoding genes all the proteins mapping to the PFHBI locus as possible and then to prioritise them for mutation screening in a PFHBI panel.

Strengths and weakness

Some of the problems experienced when using ESTs were that they are single reads and therefore the probability of sequencing errors are high. It has also been found that ESTs are contaminated with intronic sequences (Wolfsberg and Landsman, 1997). The problem with the EST databases, LLNL site and GENEMAP99 used in this study, was the lack of updating of ESTs mapping to the PFHBI locus, during 1999 to 2003. Another shortcoming of EST databases is that some genes will not be well represented because they are poorly expressed in some tissues, and, of course the contrary is true for highly expressed genes.

Both STACK and UNIGENE incorporate alternatively spliced transcripts into their clustering algorithm (Bouck *et al.*, 1999) and since it has been hypothesised that the cause of PFHBI might be at the developmental level, and alternatively spliced transcripts have been implicated in the regulatory mechanisms of development (Faustino and Cooper, 2003), it was important to include alternatively spliced transcripts in the PFHBI-

causative gene search. A shortcoming of both STACK and UNIGENE is that ESTs which are poorly expressed will be missed and thus not incorporated into their databases.

As mentioned later, gene prediction programmes have their strengths and weaknesses too, it is therefore important to use as many programmes as possible when identifying genes computationally.

4.5 AN INTEGRATED MAP OF THE PFHBI LOCUS

Since the information relevant to the PFHBI locus is scattered over a number of databases, namely, Ensembl, LLNL, NCBI, it was important to integrate all the physical and genetic data into a composite map (fig. 3.15). It facilitated navigation around the physical landscape of the PFHBI locus. In addition, the map had to be updated regularly as the sequencing data of chromosome 19 was constantly changing.

When integrating the physical map, a number of discrepancies were found; these included, wrongly assigned ESTs. For example, a number of ESTs were assigned to the wrong chromosome (table 3.6) and a LLNL clone assigned to the PFHBI locus mapped to two different positions on contig NT_011109 (fig. 4.1). The latter discrepancy may be due to gene duplication at this position of chromosome 19. Katsanis *et al.*, (2001) encountered similar problems during his study of the evaluation of the quality of the draft human genome sequence using experimental and computational methods.

In this study, the bioinformatic findings supported the experimental results, except in the case of *D19S866*. It was found that the sequence of the genetic marker *D19S866* has BLAST homology to clones AC008655 and AC020906 (fig. 4.3), whereas on amplification of these clones a product was only obtained with clone AC020906. Possible explanations for the discrepancy between the BLAST and the experimental data are that the insert may have been lost during the isolation procedure, or a PCR amplification problem was encountered or the wrong clone was used. Subsequently, *D19S866* was

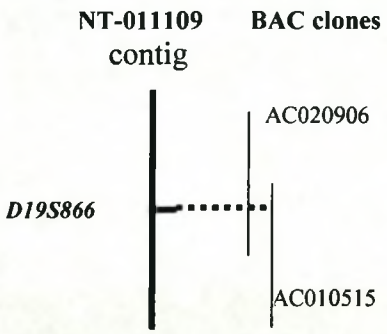


Figure 4.3 Alignment of BAC clones AC020906 and AC010515 with marker *D19S866* on contig NT_011109 using a bioinformatic approach.

mapped bioinformatically to the contig NT_011109 and could be positioned on the integrated map of the PFHBI locus (fig. 3.15).

Gaps in contig NT_011109

In May 2003, using "old" contig data, nine gaps were found on the map generated, (table 3.10). Gaps can arise when overlaps between end clones are too small to be detected by

"fingerprints" (and therefore will only be detected once the end clones are sequenced), or where duplicated segments are wrongly assigned, or in GC-rich regions of the genome, which are difficult to sequence (IHGSC, 2001) or when no overlapping fragments exist. The nine gaps found in May 2003 when aligning the "old" BAC contig against the most recent contig NT_011109 (table 3.10) have subsequently been filled, except for two clone gaps; however, the position of these two gaps have not been published (Grimwood *et al.*, 2004) and they may lie in the PFHBI area. It was not possible to detect these gaps in the current study.

An earlier map of the PFHBI locus

In 2001, Christoffels also generated a map of the PFHBI locus using a bioinformatics approach, as part of his doctoral thesis (Christoffels, 2001). He used draft sequence, generated from the three earlier contigs NT_011190, NT_011140 and NT_011157 (fig. 3.13), which was released in June 1999 by JGI. In his study, he identified seven positional PFHBI candidate genes, namely, *KPTN, GLTSRC2, NUCB1, TEAD2, CD37, DKF2P761A179* and *CGI-123*, all shown to be expressed in heart tissue. The major difference between the maps of Christoffels and the current study are that Christoffels (2001) used earlier draft sequence from the three contigs detailed above, whereas in the present study, sequence data from the newly merged contig NT_011109 was used. In addition, the order of genes and markers has also changed on the map generated in this study (fig. 3.15) (May 5, 2003).

Genes at the PFHBI locus

Predicted genes

As an outcome of the HGP, a total of 82 genes were placed at the PFHBI locus, of which 35 are annotated and 47 are predicted genes (tables 3.11 and 3.12). Included in the list of 47 predicted genes are three genes, namely, *LOC147891*, *FLJ20401* and *FLJ20643*, which were generated from the earlier EST searches (table 3.5) (section 3.2).

Only a limited amount of information was present in publicly available databases concerning the functional domains of any of the predicted proteins encoded by the 47 predicted genes. Based on the available information summarised in table 3.12 (section 3.4.1), predicted genes generally received a lower priority rating as possible candidate genes for PFHBI. Of these, four predicted genes, namely, *LOC147891*, *LOC16307, LOC147872* and *LOC95515* (fig. 3.15), were considered candidates which might be worthwhile pursuing, based on their signature domains. The predicted gene *LOC95515* is of particular interest, since its predicted sequence places it in a family of developmentally regulated mammalian transcription factors and it has often been suggested that a mutation occurring in a gene involved in cardiac development (see section 1.2.3) (Basson *et al.*, 1999) may cause PFHBI. The reasons for not screening predicted genes in this study are discussed in more detail below. Nonetheless, one cannot exclude the remaining predicted genes from being plausible PFHBI disease-causative candidate genes.

However, there are number of pitfalls and problems associated with gene prediction programmes (Mathé *et al.*, 2002), which are related to the complexity of eukaryotic genes, which must be taken into account when interpreting gene prediction data. One of the problems encountered is that of very long genes, for example, *DMK* is composed of 79 exons spanning 2.3Mb (Nobile *et al.*, 1997). The exons are interspersed with extremely long introns, some greater in length than 100kb, which make up more than 99% of the gene. Yet in some other genes, although these are rare causes, exons span only 3bp, and are easily missed if they are bordered by long introns (Nobile *et al.*, 1997). Database sequences often contain errors, although in this study sequences of predicted genes were not analysed, sequence errors were noted within annotated genes, namely, *BAX* and *GSY1* (section 3.6.2, tables 3.14 and 3.15) (Katsanis *et al.*, 2001). Thus, predicted genes resulting from gene prediction programmes should be regarded with caution, until the "virtual" proteins are verified biologically, as an active protein (Mathé *et al.*, 2002). However, gene prediction algorithms are improving constantly and more reliable data are generated at a rapid rate.

Annotated genes

During this study, 35 annotated genes were placed on contig NT_011109 at the PFHBI locus (section 3.4.1), of which 26 genes were reported to be expressed in cardiac tissue (table 3.11), thereby strengthening their candidacy for involvement in a cardiac disorder. However, there are also a number of other genes which, although not reported to be expressed in heart, based on their functions cannot be excluded from the list of candidates

to be screened, namely, *BAX*, *KIR2.4* and *CAPB5*. The expression profiles of a number of annotated genes are incomplete, since the UNIGENE database only contains expression data of cDNA sequences of tissues from which cDNA libraries have been constructed, though now there is more information on tissue and developmental stage expression obtained from microarray data but the data are yet to be integrated into annotated genome sequence sites like ENSEMBL and UCSC. Based on this, one cannot exclude non-cardiac expressed genes as plausible candidates and therefore *BAX* and *KIR2.4* were screened in the present study. During this study, *TEAD2* was not screened for PFHBI mutations because its exonic and intronic organisation was not characterised. Subsequently, new information on its genetic structure has been deposited into the Ensembl database, making it an attractive candidate gene for future mutation screening.

Previously, ESTs were randomly sequenced and deposited into the databases. However, attempts have been made to overcome this limitation, for example, a cardiovascular expression database has been established which gives information about the expression profile of genes during different stages of heart development (Liew *et al.*, 1994; Hwang *et al.*, 1997). This resource may prove useful in the PFHBI gene search, as expression of the causative gene may be developmentally regulated. There is an urgent need to increase the cardiovascular expression data in the EST database because Dempsey *et al.*, (2000) reported that the cardiac ESTs comprises of less than 4% of the total number of ESTs in Genbank (section 1.4.3.1). If this situation continues, it may hamper the chances of finding the PFHBI gene using EST data.

4.6 PFHBI CANDIDATE GENES

Genes provisionally excluded in the present study

Position-independent genes

p115-RhoGEF

In this grouping of genes, using a labeled *p115-RhoGEF* cDNA probe, the cognate gene was shown to map to two chromosome 19 radiation hybrids but not to any of the DNA clones spanning the PFHBI locus (fig. 3.18). In addition, no *RhoGEF* sequence was found at the PFHBI locus when a BLAST2 search of *p115-RhoGEF* sequence with each DNA clone was performed. As both chromosome 19 radiation hybrids spanned chromosome 19q13.1 to q13.3, it appeared, at the time, that the gene either lay outside the PFHBI locus or in a gap within the contig spanning the locus. The former assumption was confirmed when further sequencing data became available, and a bioinformatic search placed the gene centromeric to the PFHBI locus (fig. 3.19), thus excluding the *p115-RhoGEF* gene as a PFHBI-causative gene.

Cx

Although a *Cx* PCR product was not generated from PFHBI-spanning clones, using *Cx* degenerate primers, a product was obtained with chromosome 19 radiation hybrids which encompass the PFHBI locus and the flanking regions of chromosome 19q13.1 to 13.3 (section 1.4.2.1, fig. 1.11). This data suggested that a *Cx* might be present, either in a gap within the contig spanning the PFHBI locus (Grimwood *et al.,* 2004), or outside the PFHBI locus. A recent bioinformatic search found no *Cx*-like gene when the sequence was BLAST-searched against the finished chromosome 19 contig (2nd Sept, 2004).

However, a homology search of the translated PCR product against the protein database revealed an 83% homology to the protein GJA4 on chromosome 1p35.1, which is encoded by *Cx37* (figs. 3.16 and 3.17). If a duplicate *Cx37* was present in the region, then the homology search against contig NT_011109 should show it, thus it appears that a *Cx* is not present at the PFHBI locus. An explanation might be that the product generated from the radiation hybrid may very well be a hamster *Cx*, since the two somatic cell hybrids, ORIM7-1 and 1219G2, were produced from the fusion of hamster A3 cells and parental lymphoblastoid cells carrying containing chromosome 19q13.1 to q13.3 (Schonk *et al.*, 1989). Alternatively, a *Cx* may lie in one of the remaining unsequenced gaps of chromosome 19 as discussed earlier (Grimwood *et al.*, 2004).

CTG repeat expansions

Van der Merwe *et al.*, (1988) and de Meeus *et al.*, (1995) observed anticipation in PFHBI-affected and CCD-affected individuals, respectively, (section 1.3.2.3.4) and since unusual triplet repeat expansions have been associated with a number of diseases showing anticipation (see section 1.4.1.3) (Sheffield *et al.*, 1995; Rosenberg, 1996; Liquori *et al.*, 2001), the search for an expanded triplet repeat was undertaken.

The reasons for investigating the possibility of a CTG expansion in particular as causing PFHBI, rather than any of the other pathological triplet repeats, was that DM1, a multisystemic disease often associated cardiac conduction abnormalities, is caused by an expansion of a CTG repeat within a gene-encoding a serine/threonine protein kinase

(Aslanidis *et al.*, 1992; Buxton *et al.*, 1992). Since *DMK* also maps to chromosome 19, within 2.1Mb of the PFHBI centromeric marker *D19S902,* it was reasoned that genes with a CTG repeat expansion might be involved in the pathology of both diseases. The rationale was that because chromosome 19 shows evidence of genomic duplication with 7.35% of its sequence sharing sequence homology to more than one location in the genome (Grimwood *et al.*, 2004), a duplicate *DMK* might cause PFHBI. Furthermore, within the human genome paralogous genes occurring within 50kb or greater of DNA are not uncommon (Strachan and Read, 2000).

However, screening of the entire genome of selected PFHBI-affected individuals (fig. 2.1) by the Schalling group revealed no disease-causing CTG repeat expansions (fig. 3.20) (personal communication, Schalling). Since no CTG expansions were found, a search for other triplet or tetranucleotide repeat sequences should be investigated. Given that the entire sequence of the PFHBI locus is available in public databases, a modern alternative to the RED technique could be to download and to scan the PFHBI target sequence for repeat sequences using a repeat finder programme. Identified repeat sequences could then be tested experimentally for pathological expansions in PFHBI-affected individuals. This would accelerate the search for possible pathological expansions and would target only specific repeat motifs found at the PFHBI locus.

In fact, this bioinformatic approach, followed by PCR amplification of the identified repeat motif, has already been used by PFHBI team members. For example, February (2002) screened the CTG repeat of *NUCB1.* Furthermore, the GAG and GAT repeat

motifs of *HRC* had already been described in the literature (Hofmann *et al.*, 1991) and were therefore analysed by PCR-amplification for unusual expansions by de Meeus *et al.*, (1995) and Christoffels (1997) (see the following section "Genes provisionally excluded in previous studies).

Position-dependent genes

The position-dependent candidate genes selected for mutation screening from the first three most prioritised groups were *BAX, KCNA7*, *KIR2.4*, *LIN-7B*, and *GSY1* (table 3.13).

BAX, KCNA7, KIR2.4, LIN-7B* and *GSY1

Mutation screening by direct sequencing of the exonic and flanking intronic sequences the selected genes revealed no disease-causing mutations (table 4.1). With the exception of *GSY1,* these genes can only be provisionally excluded because the remaining intronic and UTR regions should also be investigated, for possible disease-causative mutations. For example, in dilated cardiomyopathy, mutations were found in the regulatory region and the coding regions of the human cardiotrophin gene (*CTF1*) (Erdmann *et al.*, 2000).

Exclusion of previously identified candidates

As part of this study, genetic fine mapping of the PFHBI locus has excluded four of the seven candidates identified by Christoffels (2001), namely, *KPTN, GLTSRC2*, *CGI-123* and *DKF2P761A179* (fig. 4.4). Another PFHBI team member screened and subsequently provisionally excluded *NUCB1* as a PFHBI disease-causative gene (F. February, 2002).

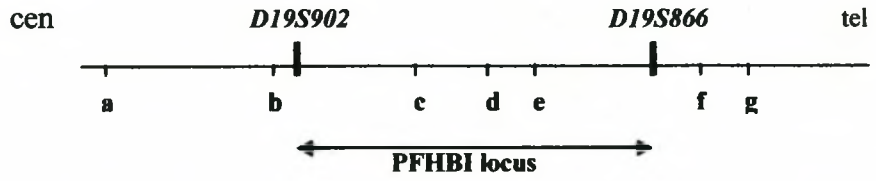


Figure 4.4 Positions of previously identified PFHBI candidate genes on the current PFHBI map. The genetic markers *D19S902* and *D19S866* flanking the PFHBI locus. cen, centomeric end; tel, telomeric end; a, *KPTN*; b, *GLTSRC2*; c, *NUCB1*; d, *TEAD2*; e, *CD37*; f, *CGI-123*; g, DKFZp761A179.

The remaining two candidates, *TEAD2* and *CD37,* prioritised by Christoffels, map to the PFHBI locus and, as yet, have not been screened. The gene *CD37* was also prioritised in the present study, and by Brink (1997), as a plausible PFHBI candidate gene (section 3.4.2).

Genes provisionally excluded in previous studies

As part of an M.Sc study, Christoffels (1997) investigated two triplet repeats GAG and GAT within exon 1 of *HRC* which were of interest, because of the anticipation reported by Van der Merwe *et al.*, (1988). In the general Caucasian population the $(GAT)_n$ repeat motif comprises 12 to 15 repeats and is tetra-allelic, whereas the $(GAG)_n$ repeat motif comprises 8 or 9 repeats and is biallelic (Christoffels, 1997). His investigation showed no expansions of either of these triplets in PFHBI-affected individuals. Following the same rationale, De Meeus *et al.*, (1995) screened the same two triplet repeats in the CCD-affected Lebanese kindred but again no abnormal repeat expansions were found. Subsequently, Dr Black sequenced *HRC* in its entirety but found no disease-causing

mutation (personal communication, Black). Based on all these investigations, *HRC* can now be excluded as PFHBI-causative.

Table 4.1 Excluded and provisionally excluded PFHBI genes

Genes screened	5'UTR	3'UTR	Exons screened	Introns screened	Approach used
Present study					
P115-RhoGEF					experimental
Cx					experimental
CTG expansion					RED analysis
BAX	nd	nd	1-6	nd	sequencing
KCNA7	nd	nd	1 & 2	nd	sequencing
KIR2.4	nd	nd	1-3	nd	sequencing
LIN-7B	nd	nd	1-5	nd	sequencing
GSY1	com	nd	1-3	nd	sequencing
			6 to 10	nd	sequencing
			13-16 (B to D)	nd	sequencing
KPTN					BLAST-search
GLTSRC2					BLAST-search
CGI-123					BLAST-search
DKF2P761A179					BLAST-search
Previous studies					
[a]*HRC*	com	com	1-6	com	sequencing
[b]*NUCB1*	nd	nd	1-13	nd	sequencing
[c]*SNRP70*	nd	nd	1-10	nd	sequencing & PCR-SSCP
[c]*SR-A1*	nd	nd	6	nd	sequencing
[d]*GSY1*	nd	nd	4, 5	nd	PCR-SSCP
			11, 12	nd	PCR-SSCP

BAX, BCL2-associated X protein; *KCNA7,* potassium voltage-gated channel shaker-related subfamily member 7; *KIR2.4,* potassium inwardly-rectifying channel, subfamily J, member 14; *LIN-7B,* lin-7 homolog B (*C. elegans*); *GSY1,* glycogen synthase 1;*KPTN*, actin-binding protein; *GLTSRC2,* unknown mRNA ; *CGI-123,* CGI protein;*DKF2P761A17,* unknown mRNA; *HRC,* histidine-rich calcium binding protein; *NUCB1,* nucleobindin 1; *SNRP70,* small nuclear ribonucleoprotein 70kDa polypeptide (RNP antigen); *SR-A1,* serine arginine-rich pre-mRNA splicing factor SR-A1 ; nd, not done, com, completed; a, personal communications, Black; b, February, 2002; c, Du Plessis, 2004; d, Makubalo, 2000.

Makubalo (2000) screened four exons of *GSY1,* which contained domains conserved across species, for mutations using PCR-SSCP (section 1.5.4). The rationale was that any sequence variation occurring in these conserved regions might have a deleterious effect on the function of the protein and possibly cause PFHBI. As part of the present study, the remaining exons and 5'UTR regions of *GSY1* were screened (section 3.6.2). No PFHBI-causative mutations were found.

In a study by F. February (2002), *NUCB1* was screened for PFHBI-causative mutations and excluded. In a recent study by Du Plessis (2004), the role of mutations in genes implicated in cardiac development was investigated. The delicate control of alternative splicing is thought to be pivotal in development (Lopez, 1998) and, therefore, any imbalance in this mechanism may be disease-causative. The gene *SNRP70* and two *SR-A1*-encoding domains located in exon 6 of the *SR-A1* gene were selected for mutation screening, since both genes map to the PFHBI locus, are splicing factors and are expressed in cardiac tissue. Furthermore the *SR-A1*-encoding domains play a role in the binding of the splicing machinery. PCR-SSCP mutation screening of the nine exons, the flanking intronic sequences of *SNRP70* and the two *SR-A1*-encoding domains, namely, *SR* (serine, arginine) and the *DR* (aspartic acid, arginine) of *SR-A1* did not reveal any disease-causative mutations, thus excluding *SNRP70* and exon 6 of *SR-A1* as PFHBI–causative.

In summary, only one of nine genes investigated using direct sequencing, namely, *HRC,* can be excluded because the entire gene was investigated for disease-causative mutations. The remaining eight genes (table 4.1) are only provisionally excluded since regulatory and intronic and regions should be screened for disease-causative mutations. Even if the gene is completely excluded, in the case of more compelling candidates, further investigation at the mRNA and protein level might be warranted.

4.7 LIMITATIONS OF THE PRESENT STUDY AND FUTURE STUDIES

The occurrence of BB disorders similar to PFHBI is not restricted to South Africa and is reported globally (table 1.1, section 1.3.2.3). It is proposed that the discovery of the PFHBI gene and its causative mutation will assist in an understanding of the genetic mechanisms underlying these cardiac conduction disorders. It is envisaged that combined knowledge will eventually translate into improved diagnosis and management of the disease.

Complete gene mutation screening

Many familial disorders have been associated with point mutations or deletions in exonic or in the flanking intronic sequences of genes. For example, in ICCD a mutation occurs in exon 12 of *SCN5A*, resulting in a substitution of cysteine 514 for glycine (G514C) in the channel protein (Tan *et al.*, 2001). In PCCD, again it is a mutation in *SCN5A* that is responsible for the disease, but it occurs in the highly conserved donor splice site of

intron 22, which results in an abnormal transcript and consequent in frame skipping of exon 22 (Schott *et al.*, 1999). However, mutations can also occur in either the regulatory or intronic regions of a gene. For example, in dilated cardiomyopathy, mutations were found in the regulatory region and the coding regions of the human cardiotrophin gene (*CTF1*) (Erdmann *et al.*, 2000). For this reason, one cannot exclude the possibility of the PFHBI-causing mutation occurring in the non-coding or regulatory regions of the responsible gene. Hence, in the present study, one of the limitations is that, with the exceptions of *HRC* in which the whole gene was sequenced, and *GSY1* in which the 5' UTR region was screened, only the exonic and the flanking intronic regions of the PFHBI candidate genes were screened for mutations.

Ascertainment bias

One question raised during this study is how real is the phenomenon of anticipation in PFHBI? The findings of Van der Merwe and colleagues (1988) in a South African family and De Meeus and colleagues (1995) in a Lebanese family both showed evidence of anticipation as a feature of the disorder. Since, in both cases, the families were known to the investigating clinicians, the family members had regular examinations and were closely monitored. Therefore, it can be argued that these studies might suffer from ascertainment bias.

Comparative genomics

Since a number of predicted genes have been placed at the PFHBI locus, comparative genomics, i.e., comparing human sequences to sequences from other organisms, for example, *M. musculus*, *D. rerio*, and *Takifugu*, is a useful resource to support the existence of these genes and can help to annotate them (Nobregia and Pennacchio, 2004). For example, Du Plessis (2004) used murine sequences to find homologous genes in humans, which had not yet been mapped to the PFHBI locus but were on the syntenic region of mouse chromosome 7. Using this approach, he identified two novel genes at the PFHBI locus, which showed similarity to the genes encoding murine small ubiquitin-related modifier-1 (*sumo-1*) (Shao *et al.*, 2004) and cytochrome oxidase VIC (*cox6c*) (Bachman *et al.*, 1997) and on assessment, the former was considered a plausible candidate gene for PFHBI. Both these predicted genes have been placed bioinformatically on the integrated physical and genetic map (fig. 3.15). The first gene participates in the apoptotic pathway and is now known to tag proteins destined for degradation in the heart and brain (Shao *et al.*, 2004), however, not enough information was available from the database to explore the human homolog. The second gene is involved in the electron transport chain which results in energy production and a mutation in this gene would have a multisystemic effect.

The mouse genome has already been particularly valuable in the search for homologous genes in the human genome, since 82%, or 45Mb, of chromosome 19 is orthologous to mouse chromosome 7 (Grimwood *et al.*, 2004). For example, as part of this study, the

coding sequence and the genomic organisation of *KCNA7* (AF315818) (fig. 3.21) (Bardien-Kruger *et al.*, 2002) were determined using the alignment of a human BAC clone AC0008687 with published mouse cDNA sequence (AF032099) (Kalman *et al.*, 1998).

In addition, animal models of human disease, e.g., mouse and zebrafish (Sehnert and Stainier, 2002) also play an important role in the functional analysis of genes; understanding the manner in which specific mutations affect the heart in these organisms will help in an understanding of the functioning human orthologs and their possible role in cardiac conduction diseases, as discussed for DM (see section 1.2.3.1) (Mathé, 2002).

Gene density on chromosome 19

A number of researchers working with this region of chromosome 19 appear to be experiencing difficulties in finding "their" disease-causative gene for disorders, for example, LGMD2 (Driss *et al.*, 2000), CED (Ghadami *et al.*, 2000) and DFNA4 (Chen *et al.*, 1995) (section 1.4.4, fig. 1.14). This may largely be due to the nature of chromosome 19 sequence (Grimwood *et al.*, 2004), namely, the gene density, which results in the large number of plausible candidate genes present at different loci, e.g., the estimated number of annotated genes at the PFHBI locus is 63. Other factors hampering all gene searches on chromosome 19 are probably due to the high G+C content which interferes with sequencing. Additionally, numerous CpG islands and the density of repetitive DNA could

cause problems when applying gene prediction programmes to sequence data from this area, which can result in a number of false positives (Mathé, 2002).

A comparison of the status of PFHBI with other gene searches

In most cases, gene searches are part of a team effort (section 1.6). The searches for FRDA and RP13 genes have been successful, whereas the searches for the PFHBI and the KWE genes, in the latter case as far as we know, are still ongoing. There are a number of similarities between the PFHBI and the KWE gene searches (section 1.6.2). At the start of each study, linkage analysis and fine mapping reduced the PFHBI (Brink *et al.*, 1995) and the KWE (Starfield *et al.*, 1997) loci to 10cM, subsequently these loci have been reduced, to a physical distance of 2.4Mb for PFHBI and 635,404bp for KWE (Appel *et al.*, 2002). In addition, both groups generated a detailed physical and transcript map of each locus, e.g., with the PFHBI study, a computational method was used and with KWE, the BAC contig was sequenced and then bioinformatics were applied (Appel *et al.*, 2002). Ideally, the reduction of the PFHBI locus to a region comparable in size to the KWE locus is desirable but, even then, identifying the disease-causing gene and the associated mutation may prove to be challenging, as it has been for KWE.

Mapping data obtained from more than one family, as was the case for the successful RP13 gene search, is useful, because the chances of meiotic events are increased, thus increasing the chance of reduction of the disease-causing locus. For example, in the RP13 gene search, three families were investigated, namely, a large SA family of British

ancestry, a family of British descent living in the USA and a British family based in Britain (section 1.6.1). In the PFHBI search, there was an initial collaboration with a French group from the Inserm laboratory, Paris, a collaboration which proved to be useful to both groups, the French group's search locus was larger, 13cM in length (de Meeus, *et al.*, 1995) compared to the PFHBI locus, which had already been reduced to 7cM (fig. 1.9) (personal communication, De Jager, 1998). When the two groups met, based on combined data, the French group reduced their locus to 7cM. As part of the present study, the most centromeric marker of the CCD locus i.e., *D19S606*, (fig. 1.9) was typed in the PFHBI families, which indeed reduced the locus to 4cM and confirmed the French group's data excluding the proximal regions. The collaboration has subsequently lapsed and it has not been possible unable to establish what stage they has been in CCD gene search, since no further publication or communication has been forthcoming.

In addition to genomic sequencing, other experimental methodologies could be used to strengthen the candidature of attractive genes. Ideally, examining expression profiles of the genes in the conduction tissue by RT PCR or by raising antibodies to investigate the distribution of the protein products would be valuable. However, a biopsy of human cardiac tissue presents challenges. An alternative approach may be to use the bovine heart, in which conduction tissue is visible as what is called the ox "false tendon" (Obbiassi *et al.*, 1987).

Thus it is hoped that, although there are apparently no other non-founder PFHBI families in South Africa or worldwide, recruitment of new members of the existing families will identify new meiotic recombination breakpoints which will further reduce the number of genes to be screened for disease-causing mutations. Database searches for new polymorphic markers will also be useful to further reduce the PFHBI locus.

As part of the ongoing research the suggested candidate genes, *TEAD2* and *CD37*, should be screened for PFHBI disease-causative mutations. The predicted proteins placed at the PFHBI locus should be re-examined since a number of these genes are either removed or replaced by annotated genes. As new genes are placed at the PFHBI locus, the genes listed for screening should be re-prioritised as PFHBI candidates. In addition, the integrated map of the PFHBI locus should be updated on a regular basis as new data are deposited into the databases.

Ultimately, when the region is small enough, sequencing across the critical region of an affected and an unaffected PFHBI individual should be pursued to find the disease-causative mutation.

It is envisaged that the identification of PFHBI-causative gene and associated mutation will provide a platform for further studies to understand the pathophysiology, not only of PFHBI, but also of other more commonly occurring conduction disturbances.

REFERENCES

Adams MD, Kelly JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, *et al.* (1991). Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* **252**:1651-1656.

Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ (1990). Basic Local Alignment Search Tool. *Proc Natl Acad Sci USA* **87**:2284-2268.

Annane D, Duboc D, Mazoyer B, Merlet P, Fiorelli M, Eymard B, *et al.* (1994). Correlation between decreased myocardial glucose phosphorylation and the DNA mutation size in myotonic dystrophy. *Circulation* **90**:2639-2634.

Annane D, Merlet P, Radvanyi H, Mazoyer B, Eymard B, Fiorelli M, *et al.* (1996). Blunted coronary reserve in myotonic dystrophy: an early gene-related phenomenon. *Circulation* **94**:973-977.

Appel S, Filter M, Reis A, Hennies HC, Bergheim A, Ogilvie E, *et al.* (2002). Physical and transcriptional map of the critical region for keratolytic winter erythema (KWE) on chromosome 8p22-p23 between *D8S550* and *D8S1759*. *Eur J Hum Genet* **10**: 17-25.

Apte SS, Mattei MG and Olsen BR (1995). Mapping of the human *BAX* gene to chromosome 19q13.3-q13.4 and isolation of a novel alternatively spliced transcript, *BAX* delta. *Genomics* **26**:592-594.

Ashworth LK, Batzer MA, Brandiff B, Branscomb E, De Jong P, Garcia E, *et al.* (1995). An integrated metric physical map of human chromosome 19. *Nat Genet* **11:**422-427.

Aslanidis C, Jansen G and Amemiya C (1992). Cloning of the essential dystrophy region and mapping of the putative defect. *Nature* **355**:548-551.

Bachinski LL, Krahe R, White BF, Wieringa B, Shaw D, Korneluk R, *et al.* (1993). An informative panel of somatic hybrids for physical mapping on human chromosome 19q. *Hum Genet* **52**:375-387.

Bachman NJ, Riggs PK, Siddiqui N, Makris GJ, Womack JE and Lomax (1997). Structure of the human gene (*COX6A2*) for heart/muscle isoform of cytochrome c oxidase subunit VIa and its chromosomal location in humans, mice, and cattle. *Genomics* **42**:146-151.

Bai G, Zhang Z, Werner R, Nuttall FQ, Tan AWH and Lee EYC (1990). The primary structure of rat liver glycogen synthase deduced by cDNA cloning. *J Biol Chem* **265**:7843-7848.

Bardien-Kruger S, Wulff H, Arieff Z, Brink P, Chandy KG and Corfield V (2002). Characterisation of the human voltage-gated potassium channel gene, *KCNA7*, a candidate gene for inherited cardiac disorders and its exclusion as cause of progressive familial heart block I (PFHBI). *Eur J Hum Genet* **10**:36-43.

Barinaga M (1998). Tracking down mutations that can stop the heart. *Science* **281**:32-35.

Basson CT, Cowley GS, Solomon SD, Weissman B, Poznanski AK, Traill TA, Seidman JG and Seidman CE (1994). The clinical and genetic spectrum of the Holt-Oram syndrome (heart-hand syndrome). *New Engl J Med* **330**:885-891.

Basson CT, Solomon SD, Weissman B, MacRae CA, Poznanski AK, Prieto F, *et al.* (1995) Genetic heterogeneity of heart-hand syndromes. *Circulation* **91**:1326-1329.

Basson CT, Bachinsky DR, Lin RC, Levi T, Elkins JA, Soults J, *et al.* (1997). Mutations in human TBX5 cause limb and cardiac malformation in Holt-Oram syndrome. *Nat Genet* **15**:30-35.

Basson CT, Huang JY, Lin R, Bachinsky DR, Weremowicz S, Vaglio A, *et al.* (1999). Different *TBX5* interactions in heart and limb defined by Holt-Oram syndrome mutations. *Proc Natl Acad Sci USA* **96**:2219-2924.

Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, Eddy SR, *et al.* (2002). The Pfam Protein families database. *Nucleic Acids Res* **30**:276-280.

Beckmann JS and Weber JL (1992). Survey of human and rat microsatellites. *Genomics* **12**:627-631.

Benson DA, Boguski M, Lipman DJ and Ostell J (1996). Genbank. *Nucleic Acids Res* **24**:1-5.

Benson G (1999). Tandem repeats finder: program to analyse DNA sequences. *Nucleic Acids Res* **27**:573-580.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J and Wheeler DL (2003). GenBank. *Nucleic Acids Res* **31**:23-7.

Boguski MS, Lowe TM and Tolstoshev CM (1993). dbEST-database for "expressed sequence tags". *Nat Genet* **29**:234-238.

Boguski M and Schuler GD (1995). Establishing a human transcript map. *Nat Genet* **10**:369-371.

Botstein D, White RL, Skolnick MK and Davis RW (1980). Construction of a genetic linkage map using restriction fragment length polymorphisms. *Am J Hum Genet* **32**:314-331.

Bouck J, Yu W and Worley K (1999). Comparison of gene indexing databases. *Trends Genet*. **4:**159-162.

Brandriff BF, Gordon LA, Fertitta A, Olsen AS, Christensen M, Ashworth LK, *et al.* (1994). Human chromosome 19p: A fluorescence-based restriction map with genomic distance estimates for 79 intervals spanning 20Mb. *Genomics* **23**:582-591.

Brink A J and Torrington M (1977). Progressive familial heart block - two types. *S Afr Med J* **52**:53-59.

Brink PA, Moolman JC, Ferreira A, De Jager T, Weymar HW, Martell, Torrington M; Van der Merwe P-L and Corfield VA (1994). Genetic linkage studies of progressive familial heart block, a cardiac conduction disorder. *S Afr J Sci* **90**:236-240.

Brink PA, Ferreira A, Moolman JC, Weymar HW, Van der Merwe P-L and Corfield VA (1995). Gene for progressive familial heart block type I maps to chromosome 19q13. *Circulation* **91**:1633-1640.

Brink PA (1997). A molecular genetic approach to the aetiology of cardiovascular disease in South Africa with reference to progressive familial heart block. Doctoral thesis in Medical Biochemistry, University of Stellenbosch.

Brook JD, Shaw DJ, Thomas NST, Meridith AL, Cowell J and Harper PS (1986). Mapping genetic markers on human chromosome 19 using subchromosomal fragments in somatic cell hybrids. *Cyto Cell Genet* **41**:30-37.

Brook JD, McCurrach ME, Harley HG, Buckler AJ, Church D, Aburatani H, *et al.* (1992). Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell* **68**:799-808.

Brown SM (2000). *Bioinformatics*: *A biologist's guide to biocomputing and the internet.* Eaton Publishing, New York, USA.

Browner MF, Nakano K, Bang AG and Fletterick RJ (1989). Human muscle glycogen synthase cDNA sequence: a negatively charged protein with an asymmetrical charge distribution. *Proc Natl Acad Sci USA* **86**:1443-1447.

Brunskill EW, Witte DP, Yutzey KE and Potter SS (2001). Novel cell lines promote the discovery of genes involved in early heart development. *Dev Biol* **235**:507-520.

Bruzzone R, White TW and Paul DL (1996). Connections with connexins: the molecular basis of direct intercellular signaling. *Eur J Biochem* **238**:1-27.

Burge C and Karlin S (1997). Prediction of complete gene structures in human genomic DNA. *J Mol Biol* **268**:78-94.

Burton EA and Davis KE (2002). Muscular Dystrophy - Reason for Optimism? *Cell* **108**: 5-8.

Buxton J, Shelbourne P and Davies J (1992). Detection of an unstable fragment of DNA specific to individuals with myotonic dystrophy. *Nature* **355**:547-548.

Campuzano V, Montermini L and Molto MD (1996). Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science* **271**:1423-1427.

Chakraborty R, Kimmel M, Stivers DN, Davison LJ and Deka R (1997). Relative mutation rates at di-, tri- and tetranucleotide microsatellite loci. *Proc Natl Acad Sci USA* **94**:1041-1046.

Chandy G and Gutman GA (1995). *Voltage-gated K+ channels genes. Ligand voltage-gated ion channels. Handbook of receptors and channels.* CRC, Florida.

Chen AH, Ni L, Fukushima K, Marietta J, O'Neill M, Coucke P, Willems P and Smith RJ (1995). Linkage of a gene for dominant non-syndromic deafness to chromosome 19. *Hum Mol Genet* **4**:1073-1076.

Chen J-N and Fishman MC (2000). Genetics of heart development. *Trends Genet* **16**:383-388.

Chen Q, Kirsch GE, Zhang D, Brugada R, Brugada J, Brugada P, *et al.* (1998). Genetic basis and molecular mechanism for idiopathic ventricular fibrillation. *Nature* **392**:293-296.

Cheng G, Litchenberg WH, Cole GJ, Mikawa T, Thompson RP and Gourdie RG (1999). Development of the cardiac conduction system involves recruitment within a multipotent cardiomyogenic lineage. *Development* **126**:5041-5049.

Chou D, Miyashita T, Mohrenweiser HW, Ueki K, Kastury K, Druck T, *et al.* (1996). The *BAX* gene maps to the glioma candidate region at 19q13.3, but is not altered in human gliomas. *Cancer Genet Cytogenet* **88**:136-140.

Christoffels AG (1997). Identification of novel CA repeat markers for use in fine mapping of the PFHBI gene and screening of the *HRC* candidate gene. Master of Science thesis in Medical Biochemistry, University of Stellenbosch.

Christoffels AG, van Gelder A, Greyling G, Miller R, Hide T and Hide WA (2001). STACK: Sequence Tag Alignment and Consensus Knowledgebase. *Nucleic Acids Res* **29**:234-238.

Christoffels AG (2001). Generation of a human gene index and its application to disease candidacy. Doctoral thesis in Biochemistry, University of the Western Cape.

Clarke NR, Kelion AD, Nixon J, Hilton-Jones D and Forfar JC (2001). Does cytosine-thymine-guanine (CTG) expansion size predict cardiac events and electrocardiographic progression in myotonic dystrophy? *Heart* **86**:411-416.

Colucci WS (1996). Apoptosis in the heart. *New Engl J Med* **335**:1224-1226.

Combrink JM, Davis WH and Snyman HW (1962). Familial bundle branch block. *Am Heart J* **64**:397-400.

Cooper D, Ball EV and Krawczak M (1998). The human gene mutation database. *Nucleic Acids Res* **26**:285-7.

Cooperative Human Linkage Centre (1994). A comprehensive human linkage map with centimorgan density. *Science* **265**:2049-2054.

Corfield VA, Moolman JC, Martell R and Brink PA (1993). Polymerase chain reaction-based detection of MN blood-group specific sequences in the human genome. *Transfusion* **33**:119-124.

Coucke P, Van Camp G, Demirhan O, Kabakkaya Y, Balemans W, Van Hauwe, *et al.*, (1997). The gene for the Pendred syndrome is located between *D7S501* and *D7S692* in a 1.7cM region on chromosome 7q. *Genomics* **40**:48-54.

Coucke PJ, Van Laer L, Meyers J, Van Hawe P, Ottschytsch N, Wauters JG, Kelley P, Willems J and Van Camp G (2000). Identification of new connexin gene *GJA11* (*Cx59*) using degenerate PCR primers. *Genescreen* **1**:35.

Cotton RGH and Scriver CR (1998). Proof of "disease-causing" mutation. *Hum Mutat* **12**:1-5.

Davies MJ (1971). *Pathology of the conduction system of the heart.* Butterworths, London.

Davies MJ and Pomerance A (1972). Quantitative study of ageing changes in the human sinuatrial node and internodal tracts. *Br Heart J* **34**:150.

Davies MJ, Anderson RH and Becker AE (1983). *The conduction system of the heart.* Butterworths, London.

Davis BM, McCurrah ME, Taneja KL, Singer RH, Houseman DE (1997). Expansion of a CUG trinucleotide repeat in the 3'-prime untranslated region of myotonic dystrophy protein kinase transcript results in nuclear retention of transcripts. *Proc Natl Acad Sci USA* **128**:995.

De Jong PJ, Yokabata K, Chen C, Lohman F, Pederson L, McNinch J and van Dilla M (1989). Human chromosome-specific partial digest libraries in lambda and cosmid vectors. *Cytogenet Cell Genet* **51**:985.

De Meeus A, Stephan E, Debrus S, Jean M-K, Loiselet J, Weissenbach J, *et al.* (1995). An isolated cardiac conduction disease maps to chromosome 19q. *Circ Res* **77**:1-6.

Dempsey AA, Ton C and Liew CC (2000). A cardiovascular EST repertoire: progress and promise for understanding cardiovascular disease. *Molec Med Today* **6**:231-237.

Den Dunnen JT and Van Ommen G-JB (1999). The protein truncation test. *Hum Genet* **14**:95-102.

Donis-Keller H, Green P, Helms C, Cartinhow S, Weiffenbach B, Stephens K, *et al.* (1987). A genetic linkage map of the human genome. *Cell* **51**:319-337.

Doupnik CA, Davidson N and Lester HA (1995). The inward rectifier potassium channel family. *Curr Opin Neurobiol* **5**:268-277.

Downes GB and Gautam N (1999). The G protein subunit gene families. *Genomics* **62**:544-552.

Driss A, Amouri R, Hamida CB, Souilem S, Gouider-Khouja N, Hamida MB and Hentati F (2000). A new locus for autosomal recessive limb-girdle muscular dystrophy in a large consanguineous Tunisian family maps to chromosome 19q13.3. *Neuromusc Dis* **10**:240-246.

Dunham A, Matthews LH, Burton J, Ashurst JL, Howe KL, Ashcroft KJ, *et al.* (2004). The DNA sequence and analysis of human chromosome 13. *Nature* **428**:522-8.

Du Plessis M (2000). Development of $(CA)_n$ repeat markers in the region between genetic markers *D19S604* and *D19S866* at the PFHBI locus. B.Sc Hons in Biochemistry, University of the Western Cape.

Du Plessis M (2004). The use of the mouse regions syntenic to the PFHBI region for identifying novel genes and the mutation screening of a candidate gene(s). Master of Science thesis in Biochemistry, University of Western Cape.

Duyk GM, Kim SW, Myers RM and Cox DR (1990). Exon trapping: a genetic screen to identify candidate transcribed sequences in cloned mammalian genomic DNA. *Proc Natl Acad Sci USA* **87**: 8995-9.

Edwards A, Civittello A, Hammond HA and Caskey CT (1991). DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Hum Genet* **49**:746-756.

Edwards A, Hammond HA, Jin L, Caskey CT and Chakraborty R (1992). Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics* **12**: 241-253.

Eldstrom J, Doerksen KW, Steele DF and Fedida D (2002). N-terminal PDZ-binding domain in Kv1 potassium channels. *Febs Letters* **531**:529-537.

Erdmann J, Hassfeld S, Kallisch H, Fleck E and Regitz-Zagrosek V (2000). Genetic variants in promoter (983>T) and coding region (A92T) of the human cardiotrophin-1 gene (*CTF1*) in patients with dilated cardiomyopathy. *Hum Mutat* **16**:448.

Esscher E, Hardell L-I and Michaëlsson M (1975). Familial, isolated, complete right bundle-branch block. *Br Heart J* **37**:745-747..

Farfel Z, Bourne HR and Iiri T (1999). The expanding spectrum of G protein diseases. *New Engl Med J* **340**:1012-1020.

Faustino NA and Cooper TA (2003). Pre-mRNA splicing and human disease. *Genes and Dev* **17**:419-437.

February F (2002). Identification and mapping of expressed sequence tags, ESTs, with a view to identifying and to screening an attractive candidate gene for PFHBI. Master of Science thesis in Biochemistry, University of the Western Cape.

Fernandez P, Corfield VA and Brink PA (2004). Progressive familial heart block type II (PFHBII): a clinical profile from 1997 to 2003. *Cardiovasc J of SA* **15**:129-132.

Findlay GH, Nurse GT, Heyl T, Hull PR, Jenkins T, Klevansky H, *et al.* (1977). Keratolyic winter erythema or 'Oudtshoorn skin' a newly recognised inherited dermatosis prevalent in South Africa. *S Afr Med J* **52**:871-874.

Fishman GI and Chien KR (1997). Fashioning the vertebrate heart: earliest embryonic decisions. *Development* **124**:2099-2117.

Franco D, Lamers WH and Moorman AFM (1998). Patterns of expression in the developing myocardium: towards a morphologically integrated transcriptional model. *Cardiovasc Res* **38**:25-53.

Gastier M, Pulido JC, Sunden S, Brody T, Buetow KH, Murray JC, *et al.* (1995). Survey of trinucleotide repeats in the human genome: assessment of their utility as genetic markers. *Hum Mol Genet* **4**:1829-1836.

Ghadami M, Makita Y, Yoshida K, Nishimura G, Fukushima Y, Waikui K, *et al.* (2000). Genetics mapping of the Camurati-Engelmann disease locus to chromosome 19q13.1-q13.3. *Hum Genet* **66**:143-147.

Ghosh TK, Packham EA, Bonser EA, Robinson AJ, Cross TE and Brook JD (2001). Characterisation of the TBX5 binding site and analysis of mutations that cause Holt-Oram syndrome. *Hum Mol Genet* **10**:1983-1994.

Goldberger AL and Goldberger E (1990). *Clinical electrocardiography, a simplified approach* (4th ed). Mosby, Missouri, USA.

Goliath R, Shugart Y, Janssens P, Weissenbach J, Beighton P, Ramesar RS and Greenberg J (1995). Fine location of the locus for autosomal dominant retinitis pigmentosa on chromosome 17p. *Am J Hum Genet* **57**:962-965.

Goliath R (2000). Towards identifying the *ADRP* gene in a large South African retinitis pigmentosa family. Doctoral thesis in Human Genetics, University of Cape Town.

Gordon LA, Bergmann A, Christiansen M, Danganan L, Lee D, Ashworth LK, *et al.* (1995). A 30-Mb metric fluorescence *in situ* hybridisation map of human chromosme 19q. *Genomics* **30**:187-194.

Gorza L, Schiaffino S and Vitadello M (1988). Heart conduction system: A neural crest derivative? *Brain Res* **457**:360-366.

Gourdie RG, Mima T, Thompson RP and Mikawa T (1995). Terminal diversification of the myocyte lineage generates Purkinje fibres of the cardiac conduction system. *Development* **121**:1423-1431.

Gourdie RG, Kubalak S and Mikawa T (1999). Conducting the embryonic heart: orchestrating development of specialised cardiac tissues. *Trends Cardiovasc Med* **9**:18-26.

Graber HL, Unverferth DV, Baker PB, Ryan JM, Baba N and Wooley CF (1986). Evolution of hereditary cardiac conduction and muscle disorder: A study involving a family with six generations affected. *Circulation* **7**:21-35.

Greenberg J, Goliath R, Beighton P and Ramesar RS (1994). A new locus for autosomal dominant retinis pigmentosa on the short arm of chromosome 17. *Hum Mol Genet* **3**:915-918.

Grimwood J, Gordon LA, Olsen A, Terry A, Schmutz J, Lamerdin J, *et al.* (2004). The DNA sequence and biology of human chromosome 19. *Nature* **428**:529-535.

Groh WJ, Lowe MR. and Zipes DP (2002). Severity of cardiac conduction involvement and arrhythmias in myotonic dystrophy type 1 correlates with age and CTG repeat length. *J Cardiovasc Elect* **13**:444-448.

Grompe M (1993). The rapid detection of unknown mutations in nucleic acids. *Nat Genet* **5**:111-117.

Gros DB and Jongsma HJ (1996). Connexins in mammalian heart function. *Bioessays* **18**:719-730.

Gyapay G, Morrisette J, Vignal A, Dib C, Fizames C, Millaseau P, *et al.* (1994). The 1993-1994 Généthon human genetic linkage map. *Nat Genet* **7**:246-249.

Habets PE, Moorman AF, Clout DE, van Roon MA, Lingbeek M, van Lohuizen M, *et al.* (2002). Cooperative action of Tbx2 and Nkx2.5 inhibits ANF expression in the atrioventricular canal: implication for cardiac chamber formation. *Gene Devel* **16**:1234-1246.

Hamada H, Petroni MG, Kakunaga T, Seidman M and Stollar BD (1984). Characterisation of genomic poly(dT-dG).poly(dC-dA) sequences: Structure, organisation and conformation. *J Mol Cell Biol* **4:**2610-2621.

Hampton JR (1997). *The ECG made easy*. Livingstone Publisher, Edinburgh, UK.

Hamshere M, Cross S, Daniels M, Lennon G and Brook JD (1999). A transcript map of a 10-Mb region of chromosome 19: A source of genes for human disorders, including candidates for genes involved in asthma, heart defects, and eye development. *Genomics* **63**: 425-429.

Hardt SE and Sadoshima J (2002). A novel regulator of cardiac hypertrophy and development. *Circ Res* **90**:1.

Harper PS (1989). *Myotonic dystrophy* (2nd ed). WB Saunders, Philadelphia, USA.

Hart JH, Sharma S, elMasry N, Qui R-G, McCabe P, Polakis P and Bollag G (1996). Identification of a novel guanine nucleotide exchange factor for the rho GTPase. *J Biol Chem* **271**:25452-25458.

Hasty P, Campisi J, Hoeijimakers J, van Steeg H and Vijg J (2003). Ageing and genome maintenance: lessons from the mouse? *Science* **299**:1355-1359.

Higgins IT, Kannel WB and Dawber TR (1965). The electrocardiogram in epidemiological studies; reproducibility, validity and international comparison. *Br J Prev Med* **19**:53-68.

Hillier L, Lennon G, Becker M, Bonaldo MF, Chiapelli B, Chissoe S, *et al.* (1996). Generation and analysis of 280,000 human expressed sequence tags. *Genome Res* **6**:807-828.

Hiroi Y, Kudoh S, Monzen K, Ikeda Y, Yazaki Y, Nagai R and Komuro I (2001). Tbx5 associates with Nkx2-5 and synergistically promotes cardiomyocyte differentiation. *Nat Genet* **3**:276-80.

Hofmann SL, Topham M, Hsieh C-L and Francke U (1991). cDNA and genomic cloning of HRC, a human sarcoplasmic reticulum protein, and localisation of the gene to human chromosome 19 and mouse chromosome 7. *Genomics* **9:**656-669.

Holt M and Oram S (1960). Familial heart disease with skeletal malformations. *Br Heart J* **22**:236-242.

Howeler CJ, Busch HFM, Geraedts JPM, Nietmeijer MF and Staal A (1989). Anticipation in myotonic dystrophy: factor or fiction? *Brain* **112**:779-797.

Hulsebos T, Wieringa B, Hochstenbach R, Smeets D, Schepens J, Oerlemans F, *et al.* (1986). Toward early diagnosis of myotonic dystrophy: construction and characterization

of a somatic cell hybrid with a single human der (19) chromosome. *Cytog Cell Genet* **43**:47-56.

Hwang DM, Dempsey AA, Wang RX, Rezvani M, Barrans JD, Dai KS, *et al.* (1997). A genome-based resource for molecular cardiovascular medicine: toward a compendium of cardiovascular genes. *Circulation* **96**:4146-4203.

Huynen M, Doerks T, Eisenhaber F, Orengo C, Sunyaev S, Yuan Y, Bork P (1998). Homology-based fold predictions for Mycoplasma genitalium proteins. *J Mol Biol* **280**:323-326.

Innis MA and Gelfand DH (1990). *Optimisation of PCRs. PCR protocols: a guide to methods and applications*. Academic Press Inc, New York.

International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. *Nature* **409**:860.

Jalife J, Morley GE and Vaidya D (1999). Connexins and impulse propagation in the mouse heart. *J Cardiovasc Electrophysiol* **10**:1649-1663.

James TN (1970). Cardiac conduction system: foetal and postnatal development. *Am J Cardiol* **25**:213-226.

James TN, McKone RC and Hudspeth AS (1975). De subitaneis morbitus-familial congenital heart block. *Circulation* **51**:379-388.

James TN and Facc MD (1985). Normal variations and pathological changes in structure of the cardiac conduction system and their functional significance. *J Am Coll Cardiol* **5**:71-78.

James TN, Martin ES, Willis III PW and Lohr TO (1996). Apoptosis is a possible cause of gradual development of complete heart block and fatal arrhythmias associated with absence of the AV Node and internodal pathways. *Circulation* **93**:1424-1438.

Jansen G, Mahadevan M, Amemiya C, De Jong PJ, Amemiya C, Wormskamp NGM, *et al.* (1992). Characterisation of the myotonic dystrophy gene predicts multiple protein isoform-encoding mRNA's. *Nat Genet* **1**:261-266.

Jellinger KA and Stadelmann C (2001). Problems of cell death in neurodegeneration and Alzheimer's disease. *J Alzheimers Dis* **3**:31-40.

Jennings C (1995). How trinucleotide repeats may function. *Nature* **378**:127.

Jongsma HJ and Wilders R (2000). Gap junctions in cardiovascular disease. *Circulation* **86**:1193-1197.

Johnson DF, Hamon M and Fischel-Ghodsian N (1998). Characterisation of the human mitochondrial ribosomal S12 gene. *Genomics* **52**: 363-368.

Kahler RL, Braunwald E, Plauth WH and Morrow AG (1996). Familial congenital heart disease: familial occurrence of atrial septal defect with A-V conduction abnormalities, supravalvular aortic and pulmonic stenosis and ventricular septal defect. *Am J Med* **40**:384-399.

Kalman K, Nguyen A, Tseng-Crank J, Dukes ID, Chandy G, Hustad CM, *et al.* (1998). Genomic organisation, chromosomal localisation, tissue distribution and bio-physical characterisation of a novel mammalian Shaker-related voltage-gated potassium channel, *Kv1.7. J Biol Chem* **273**:5851-5857.

Kanai Y, Takeda O, Miura K and Kurosawa Y (1993). Novel autoimmune phenomena induced *in vivo* by a new DNA binding protein nuc: a study of MRL/n mice. *Immun Letters* **39**:83-89.

Kanzawa N, Poma CP, Takebayashi-Suzuki K, Diaz KG, Layliev J and Mikawa T (2002). Competency of embryonic cardiomyocytes to undergo Purkinje fibre differentiation is regulated by endothelin receptor expression. *Development* **129**:3185-3194.

Kass S, McRae C, Graber HL, Sparks EA, McNamara D, Boudoulas H, *et al.* (1994). A gene defect that causes conduction system disease and dilated cardiomyopathy maps to chromosome 1p1-1q1. *Nat Genet* **7**: 546-551.

Katsanis N, Worley KC and Lupski JR (2001). An evaluation of the draft human genome sequence. *Nat Genet* **29**:88-91.

Katz AM (1993). Cardiac channels. *New Engl J Med* **328**:1244-1251.

Keating M (1992). Linkage analysis and long QT syndrome. Using genetics to study cardiovascular disease. *Circulation* **85**:1973-1986.

Kelsell DP, Dunlop J, Stevens HP, Lench NJ, Liang JN, Parry G, Meuller RF and Leigh IM (1997). Connexin 26 mutations in hereditary non-syndromic sensorineural deafness. *Nature* **387**:80-83.

Kim JS, Viragh S, Moorman AF, Anderson RH, Lamers WH, *et al.* (2001). Development of the myocardium of the atrioventricular canal and the vestibular spine in the human heart. *Circ Res* **88**:395-402.

King LM and Opie LH (1998). Glucose and glycogen utilisation in myocardial ischemia-Changes in metabolism and consequences for the myocyte. *Mol Cell Biochem* **180**:3-25.

Kinoshita A, Saito T, Tomita H, Makita Y, Yoshida K, Ghadami M, *et al.* (2000). Domain-specific mutations in TGFB1 result in Camurati-Engelmann disease. *Nat Genet* **26**: 19-20.

Kirchoff S, Kim JS, Hagendorff A, Thonnissen E, Kruger O, Lamers WH and Willecke K (2000). Abnormal cardiac conduction and morphogenesis in connexin 40 and connexin 43 double-deficient mice. *Circ Res* **87**:399-405.

Kojis TL, Heinzmann C, Flodman P, Ngo JT, Sparkers RS, Spence MA, *et al.* (1996). Map refinement of locus RP13 to human chromosome 17p13.3 in a second family with autosomal dominant retinitis pigmentosa. *Am J Hum Genet* **58**:347-355.

Kroemer G (1998). Mitochondrial control of apoptosis: an overview. *Biochem Soc Symp* **66**:1-15.

Kruger O, Plum A, Kim JS, Winterhager E, Maxeiner S, Hallas G, *et al.* (2000). Defective vascular development in connexin 45-deficient mice. *Development* **127**:4179-93.

Krutovskikh V and Yamasaki H (2000). Connexin gene mutations in human genetic diseases. *Mut Res* **462**:197-207.

Kumar D, Lou H and Singal PK (2002). Oxidative stress and apoptosis in heart dysfunction. *Herz* **27**:662-668.

Kurachi Y, Jan LY and Lazdunski M (1999). *Current topics in membranes: Potassium ion channels*. Academic Press, New York.

Lanham JG, Walport MJ and Hughes GRV (1983) Congenital heart block and familial connective tissue disease. *J Rheumat* **10**:823-825.

Larsden F, Gunderson G, Lopez R and Prydz H (1992). CpG islands as gene markers for the human genome. *Genomics* **13**:1095-1107.

Lehto M, Stoffel M, Groop L, Espinosa R III, LeBeau MM and Bell GI (1993). Assignment of gene encoding glycogen synthase (GYS) to human chromosome 19 band q13.3. *Genomics* **15**:460-461.

Lenègre J (1964). Etiology and pathology of bilateral bundle branch block in relation to complete heart block. *Prog Cardiovasc Dis* **6**:409-444.

Leonoudakis D, Mailliard SW, Wingerd KL, Clegg DO and Vandenberg CA (2000). Inward rectifier potassium channel Kir2.2 is associated with Synapse-Associated Protein SAP97. *J Cell Science* **114**:987-998.

Lev M (1964). The pathology of complete atrio-ventricular block. *Progress Cardiovasc Dis* **6**:317-326.

Levin M, Johnson RL, Stern CD, Kuehn M and Tabin CA (1995). A molecular pathway determining left-right asymmetry in chick embryogenesis. *Cell* **82**:803-814.

Li QY, Newbury-Ecob RA, Terret JA, Wilson DI, Curtis ARJ, Yi CH *et al.* (1997). Holt-Oram syndrome is caused by mutations in TBX5, a member of the Brachyury (T) gene family. *Nat Genet* **15**:21-29.

Liew C-C, Hwang DM, Fung YW, Laurenssen C, Cukerman E, Tsui S and Lee CY (1994). A catalogue of genes in the cardiovascular system as identified by expressed sequence tags. *Proc Natl Acad Sci USA* **91**:10645-10649.

Liquori CL, Ricker K, Moselsey ML, Jacobsen JF, Kress W, Naylor SL, *et al.* (2001). Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of *ZNF9*. *Science* **293**:864-867.

Liu Q and Sommer SS (1994). Parameters affecting the sensitivities of dideoxy fingerprinting and SSCP. *PCR Methods Appl* **4**:97-108.

Liu XZ, Xia XJ, Xu LR, Pandya A, Liang CY, Blanton SH, *et al.* (2000). Mutations in connexin 31 underlie recessive as well as dominant non-syndromic hearing loss. *Hum Mol Genet* **9**:63-67.

Lopez AJ (1998). Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. *Ann Rev Genet* **32**:279-305.

Luderitz B (1995). *History of the disorders of cardiac rhythm.* Futura Publishing Co., New York, USA.

Mahadevan MS, Tsilfidis C, Sabourin L, Shutler G, Amemiya C, Jansen G, *et al.*, (1992). Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. *Science*. **255**:1253-1255.

Mahadevan MS, Ameniya C, Jansen G; Sabourin I, Baud S, Neville CE, *et al.* (1993). Structure and genomic sequence of the myotonic dystrophy (DM kinase) gene. *Hum Mol Genet* **2**:299-304.

Makowski L, Caspar DLD, Phillips WC and Goodenough DA (1977). Gap junction structures II. Analysis of the x-ray diffraction data. *J Cell Biol* **74**:629-645.

Makubalo Z (2000). Mutation screening of candidate genes and the development of polymorphic markers residing on chromosome 19q13.3, the progressive familial heart block I gene search area. Master of Science thesis in Medical Biochemistry, University of Stellenbosch.

Marchler-Bauer A, Panchenko AR, Shoemaker BA, Thiessen PA, Geer LY and Bryant SH (2002). CDD: a database of conserved domain alignments with links to domain three-dimensional structure. *Nucleic Acids Res* **30**:281-283.

Mandel WJ (1980). *Cardiac Arrhythmias-mechanisms, diagnosis and management*. J.B Lippincott Company, USA.

Mayall BH (1984). The DNA-based human karotype. *Cytometry* **5**:376-385.

Mathé C, Sagot M-F, Schiex T and Rouzé (2002). Current methods of gene prediction, their strengths and weaknesses. *Nucleic Acids Res* **30**:4103-4117.

McHale JC, McKie AB and Inglehearn CF (2000). Expression map of human chromosome region 17p13.3, spanning the RP13 dominant retinis pigmentosa locus, the Miller-Dieker lissencephaly syndrome (MDLS) region and a putative tumour suppressor locus. *Cytogenet Cell Genet* **88**:225-229.

McKie AB, McHale JC, Keen TJ, Tarttellin EE, Goliath R, van Lith-Verhoeven JJC, *et al.* (2001). Mutations in the pre-mRNA splicing factor gene *PRPC8* in autosomal dominant retinitis pigmentosa (RP13). *Hum Mol Genet* **10**:1555-1562.

Melacini P, Villanova C and Menegazzo E (1995). Correlation between cardiac involvement and CTG trinucleotide repeat length in myotonic dystrophy. *J Am Coll Cardiol* **25**:239-245.

Mikawa T (1999). *Cardiac lineages: In heart development*. Academic Press, San Diego, USA.

Milewski RC, Neil C, Chi C, Li J, Brown C, Lu MM, *et al.* (2003). Identification of minimal enhancer elements sufficient for Pax3 expression in neural crest and implication of Tead2 as a regulator of Pax3. *Development* **131**:829-837.

Miller RT, Christoffels AG, Gopalakrishnan c, Burke JA, Ptitsyn AA, Broveak TR and Hide WA (1999). A comprehensive approach to clustering of expressed human gene sequence: The Sequence Tag Alignment and Consensus Knowledgebase. *Genome Res* **9**:1143-1155.

Miller C (2000). An overview of the potassium channel family. *Genome Biol* **1**:1-4.

Misawa H, Kawasaki Y, Sweeney N, Jo K, Nicoll RA and Bredt DS (2001). Contrasting localisation of *MALS/LIN-7* PDZ proteins in brain and molecular compensation in knockout mice. *J Biol Chem* **276**:9264-9272.

Miquerol L, Dupays L, Theveniau-Ruissy M, Alcolea S, Jarry-Guichard T, Abran P and Gros D (2003). Gap junctional connexins in the developing mouse cardiac conduction system. *Nov Found Symp* **250**:80-279.

Modrek B, Resch A, Grasso C and Lee C (2001). Genome-wide detection of alternative splicing in expressed sequences of human genes. *Nucleic Acids Res* **29**:2850-2859.

Moolman JC (1992). A linkage study of progressive familial heart block using PCR-based detection of polymorphic loci and an investigation into the role of autoantibodies in this disease. Master of Science thesis in Medical Biochemistry, University of Stellenbosch.

Moorman AFM, De Jong F, Denyn MM and Lamers WH (1998). Development of the cardiac conduction system. *Circ Res* **82**:629-644.

Mott R (1997). EST_GENOME: a program to align spliced DNA sequences to unspliced genomic DNA. *Cabios* **13**:477-478.

Mounsey JP, Mistry D, Ai CW, Reddy S and Moorman JR (2000). Skeletal muscle sodium channel-gating in mice deficient in myotonic dystrophy protein kinase. *Hum Mol Genet* **9**:2313-2320.

Muchir A, Bonne G, van der Kooi AJ, van Meegen M, Baas F, Bolhuis PA, *et al.* (2000). Identification of mutations in the gene encoding lamins A/C in autosomal dominant limb girdle muscular dystrophy with atrioventricular conduction disturbances (LGMD1B). *Hum Mol Genet* **9**:1453-9.

Murata M, Buckett PD, Zhou J, Brunner M, Folco E and Koren G (2001). SAP97 interacts with *Kv1.5* in heterologous expression systems. *Heart Circul Physiol* **282**:2575-2584.

Myburgh DP and Steenkamp WFJ (1973). The hereditary nature of adult-onset heart block. *SA Med J* **47**:657-658.

Narula J, Haider N, and Virmani R (1996). Apoptosis in myocytes in end stage heart failure. *New Engl J Med* **335**:1182-1189.

Neitzel HA (1986). Routine method for the establishment of permanent growing lymphoblastoid cell lines. *Hum Genet* **73**:320-326.

Nerheim P, Krishnan SC, Olshansky B and Shivkumar K (2001). Apoptosis in the genesis of cardiac rhythm disorders. *Cardiol Clin* **1**:155-163.

Nguyen HH, Wolfe JT, Holmes DR and Edwards WD (1988). Pathology of the cardiac conduction system of myotonic dystrophy: a study of 12 cases. *J Am Coll Cardiol* **11**:662-671.

Nguyen-Trán VT, Kubalak SW, Minamisawa S, Fiset C, Wollert KC, Brown A B, *et al.* (2000). A novel genetic pathway for sudden cardiac death via defects in the transition between ventricular and conduction system cell lineages. *Cell* **102**:671-682.

Nievergelt CM, Smith WD, Kohlenberg JB and Schork NJ (2004). Large-scale integration of human genetic and physical maps.*Genome Res* **14**:1199-1205.

Nobile C, Marchi J, Nigro V, Robers RG and Danieli GA (1997). Exon-intron organisation of the human dystrophin gene. *Genomics* **45**:421-424.

Nobrega MA and Pennacchio LA (2004). Comparative genomic analysis as a tool for biological discovery. *J Physiol* **554**:31-9.

Nuttal FQ, Gannon MC, Kubic VL and Hoyt KJ (1994). The human liver glycogen synthase isozyme gene is located on the short arm of chromosome 12. *Genomics* **19**:404-405.

Obbiassi M, Brucato A, Meroni PL, Vismara A, Lettino M, Poloni F *et al.* (1987). Antibodies to cardiac Purkinje cells: Further characterization in autoimmune diseases and AV block. Clin Immun and Immunopath **42**:141-150.

Ohshima K, Montermini L, Wells RD and Pandolfo M (1998). Inhibitory effects of expanded GAA-TTC triplet repeats from intron I of the Friedreich ataxia gene on transcription and replication *in vivo*. *J Biol Chem* **273**:14588-14595.

Olson EN and Srivastava D (1996). Molecular pathways controlling heart development. *Science* **272**:671-676.

Olsen O, Liu H, Wade JB, Merot J and Welling PA (2002). Basolateral membrane expression of the Kir 2.3 channel is co-ordinated by PDZ interaction with Lin-7/CASK complex. *Am J Cell Physiol* **282**:C183-C195.

Oltvai ZN, Milliman CL and Korsmeyer SJ (1993). Bcl-2 heterodimerises *in vivo* with a conserved homolog, BAX, that accelerates programmed cell death. *Cell* **74**:609-619.

Oosthoek PW, Virágh S, Lamers WH and Moorman AFM (1993). Immunohistochemical delineation of the conduction system II: the atrioventricular node and Purkinje Fibres. *Circ Res* **73**:482-491.

Opie LH (1991). *The heart: Physiology and metabolism* (2nd ed). Raven Press, New York

Orho M, Nikula-Ijas P, Schalin-Jantti C, Permutt MA and Groop LC (1995). Isolated and characterisation of the human muscle glycogen synthase. *Diabetes* **44**:1099-1105.

Orita M, Suzuki Y, Sekiya T and Hayashi K (1989). Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics* **5**:874-879.

Otten AD and Tapscott SJ (1995). Triplet repeat expansion in myotonic dystrophy alters the adjacent chromatin structure. *Proc Natl Acad Sci USA* **92**:5465-5469.

Pacher P, Csordas G and Hajnoczky G (1999). Mitochondrial Ca^{2+} signaling and cardiac apoptosis. *Dev Biol* **207**:271-286.

Park M, Wu X, Golden K, Axelrod JD and Bodmer R (1996). The wingless signaling pathway is directly involved in *Drosophila* heart development. *Dev Biol* **177**:104-116.

Patten BM (1956). The development of the sino-ventricular conduction system. *Univers Mich Med Bull* **22**:1-21.

Pease W, Nordenberg A and Ladda RL (1976). Genetic counseling in familial atrial septal defect with prolonged atrio-ventricular conduction. *Circulation* **53**:759-762.

Pennisi DJ, Rentschler S, Gourdie RG, Fishman GI and Mikawa T (2002). Induction and patterning of the cardiac conduction system. *Int J Dev Biol* **46**:765-775.

Perloff JK, Stevenson WG, Roberts NK, Cabeen W and Weiss J (1984). Cardiac involvement in Myotonic Muscular Dystrophy (Steinert's Disease): a prospective study of 25 patients. *Am J Cardiol* **54**:1074-1081.

Perryman MB, Friedman DL, Fu Y-H and Caskey CT (1993). Molecular genetics of myotonic dystrophy. *Trends Cardiovasc Med* **3**:82-84.

Plaster NM, Tawil R, Tristani-Firouzi M, Canún S, Bendahou D, Tsunado A, *et al.* (2001). Mutations in *Kir2.1* cause the developmental and episodic electrical phenotypes of Anderson's syndrome. *Cell* **105**:511-519.

Pollard TD and Earnshaw WC (2002). *Cell Biology*. Saunders, USA.

Prosser J (1993). Detecting single-base mutations. Trends in Biotechnology 11:238-246.

Pusch CM, Meyer B, Kupka S, Smith RJ, Lalwani AK, Zenner HP, *et* al. (2004). Refinement of the *DFNA4* locus to a 1.44 Mb region in 19q13.33. *J Mol Med* **82**:398-402.

Ranum LPW, Rasmussen PF, Benzow KA, Koob MD and Day JW (1998). Genetic mapping of a second myotonic dystrophy locus. *Nat Genet* **19**:196-198.

Ranum LPW and Day JW (2004). Myotonic dystrophy: RNA pathogenesis comes into focus. *Am J Hum Genet* **74**:793-804.

Rosenberg RN (1996). DNA-triplet repeats and neurological disease. *New Engl J Med* **335**:1222-1224.

Rossant J (1995). Mouse mutants and cardiac development. *Circ Res* **78**:349-353.

Rudel R, Ruppersberg JP and Spittelmeister W (1989). Abnormalities of the fast sodium current in myotonic dystrophy, recessive generalised myotonia and adynamia episodica. *Muscle Nerve* **12**:281-287.

Sanchez I, Mejia RO and Friedlander RM (2001). Caspases in Huntington's disease. *Neuroscientist* **7**:480-489.

Saveliev A, Everett C, Sharpe T, Webster Z and Festenstein R (2003). DNA triplet repeats mediate heterochromatin-protein-1-sensitive variegated gene silencing. *Nature* **422**: 909-913.

Schaal SF, Seidensticker J, Goodman R, Wooley CF (1973). Familial right bundle-branch block, left axis deviation, complete heart block and early death. *Ann Int Med* **79**:63-66.

Schalling M, Hudson TJ, Buetow KH and Housman DE (1993). Direct detection of novel expanded trinucleotide repeats in the human genome. *Nat Genet* **4**:135-139.

Schonk D, Coerwinker-Driessen M, Van Dalen I, Oerlemans F, Smeets B, Schepens J, *et al.* (1989). Definition of subchromosomal intervals around the Myotonic Dystrophy gene at 19q. *Genomics* **4**:384-396.

Schott J-J, Benson DW, Basson CT, Pease W, Silberbach GM, Moak JP, *et al.* (1998). Congenital heart disease caused by mutations in the transcription factor *NKX2-5*. *Science* **281**:108-111.

Schott J-J, Alshinawi C and Kyndt F (1999). Cardiac conduction defects associate with mutations in *SCN5A*. *Nat Genet* **23**:20-21.

Schwarz-Albiez R, Dorken B, Hofmann W and Moldenhauer G (1988). The B cell-associated CD37 antigen (gp40-52). Structure and subcellular expression of an extensively glycosylated glycoprotein. *J Immunol* **140**:905-914.

Sehnert AJ and Stainier DYR (2002). A window to the heart: can zebrafish mutants help us understand heart disease in humans? *Trends Genet* **18**:491-494.

Shah ZH, Migliosi V, Miller SCM, Wang A, Friedman TB and Jacobs HT (1998). Chromosomal locations of three human nuclear genes (*RPSM12*, *TUFM*, and *AFG3L1*) specifying putative components of the mitochondrial gene expression apparatus. *Genomics* **48**: 384-388.

Shaw DJ, Brook JD, Meredith AL, Harley HG, Sarfarazi M and Harper PS (1986). Gene mapping and chromosome 19. *J Med Genet* **23**:2-10.

Shao R, Zhang FP, Tian F, Anders Friberg P, Wang X, Sjoland H and Billig H (2004). Increase of SUMO-1 expression in response to hypoxia: direct interaction with HIF-1 alpha in adult mouse brain and heart *in vivo*. *FEBs Lett*. **569**:293-300.

Sheffield VC, Weber JL, Buetow KH, Murray JC, Even DA, Wiles K, *et al.* (1995). A collection of tri- and tetranucleotide repeat markers used to generate high quality, high resolution human genome-wide linkage maps. *Hum Mol Genet* **4**:1837-1844.

Shiojima I, Komuro I, Inazawa J, Nakahori Y, Matsushita I, Abe T, Nagai R and Yazaki Y (1995). Assignment of cardiac homebox gene *CSX* to human chromosome 5q34. *Genomics* **27**:204-206.

Simon MI, Strathmann MP and Gautum N (1991) Diversity of G proteins in signal transduction. *Science* **252**:892-805.

Simon AM, Goodenough DA and Paul DL (1998). Mice lacking connexin 40 have cardiac conduction abnormalities characteristic of first-degree atrioventricular block and bundle branch block. *Curr Biol* **8**:295-298.

Slightom JL, Blechl AE and Smithies O (1980). Human fetal G gamma and A gamma-globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. *Cell* **21**:627-638.

Solovyev VV and Salamov AA (1999). INFOGENE: A database of known gene structures and predicted genes and proteins in sequences of genome sequencing projects. *Nucleic Acids Res* **27**:248-250.

Southern EM (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* **98**:503-517.

Spach MS, Heidlage JF, Dolber PC and Barr RC (2003). Electrophysiology effects of remodeling cardiac gap junctions and cell size: experimental and model studies of normal cardiac growth. *Circ Res* **86**:302-311.

Spiegel AM (1996). Defects in G protein-coupled signal transduction in human disease. *Ann Rev Physiol* **58**:143-170.

Splawski I, Shen J, Timothy KW, Lehman MH, Priori S, Robinson JL, *et al.*, (2000). Spectrum of mutations in Long-QT and syndrome genes *KVLQT1*, *HERG*, *SCN5A*, *KCNE1*, and *KCNE2*. *Circulation* **102**:1178-1185.

Srivastava D (1997). HAND Proteins: molecular mediators of cardiac development and congenital heart disease. *Trends Cardiovasc Med* **9**:11-18.

Stainier YR, Fouquet B, Chen J, Warren KS, Weinstein BM, Meiler S, *et al.* (1996). Mutations affecting the formation and function of the cardiovascular system in the zebrafish embryo. *Dev* **123**:285-292.

Stallings RL (1994). Distribution of trinucleotide macrosatellites in different categories of mammalian genomic sequence: implication for human genetic diseases. *Genomics* **21**:116-121.

Starfield M, Hennies HC, Jung M, Jenkins T, Wienker T, Hull P, *et al.* (1997). Localisation of the gene causing keratolytic winter erythema to chromosome 8p22-p23, and evidence for the founder effect in South African afrikaans-speakers. *Am J Hum Genet* **61**:370-378.

Stein L (2001). Genome annotation from sequence to biology. *Nat Rev Genet* **2**:493-503.

Strachan T and Read AP (2000). *Human Molecular Genetics* (2nd ed). BIOS Scientific publishers, Oxford, UK.

Strausberg RL, Feingold EA, Grouse LH, Derge JG, Klausner RD, Collins FS, *et al.* (2002). Generation of initial analysis of more than 15000 full length human and mouse cDNA sequences. *Proc Natl Acad Sci USA* **99**:16899-16903.

Takebayashi-Suzuki K, Paulikks LB, Eltsefon Y and Mikawa T (2001). Purkinje fibres of the avian heart express a myogenic transcription factor program distinct from cardiac and skeletal muscle. *Cardiovasc Res* **49**:417-429.

Tan HL, Bink-Boelkens MTE, Bezzina CR, Viswanathan PC, Beaufort-Krol GCM, Van Tintelen PJ, *et al.* (2001). A sodium-channel mutation causes isolated cardiac conduction disease. *Nature* **409:**1043-1047.

Tan HL, Kupershmidt S, Zhang R, Stepanovic S, Roden DM, Wilde AA, Manderson ME and Balser JR (2002). A calcium sensor in the sodium channel modulates cardiac excitability. *Nature* **415**:442-447.

Taneja KL, Lifshitz LM, Fay FS and Singer RH (1995). Poly (A) rRNA code distribution with microfilaments: evaluation by *in situ* hybridisation and quantitative digital imaging microscopy. *J Cell Biol* **128**:195.

Tapscott SJ (2000). Deconstructing myotonic dystrophy. *Science* **289**:1701-1702.

Tapscott SJ and Thornton CA (2001). Reconstructing myotonic dystrophy. *Science* **293**:816-817.

Tarttellin EE, Plant C, Weissenbach J, Bird AC, Bhattacharya SS and Inglehearn CF (1996). A new family linked to the RP13 locus for autosomal dominant retinitis pigmentosa on distal 17p. *J Med Genet* **33**:518-520.

Tautz D (1989). Hypervariability of simple sequences as a general source of polymorphic DNA markers. *Nucleic Acids Res* **17**:6463-6471.

Terrett JA, Newbury-Ecob R, Cross GS, Fenton I, Raeburn JA, Young ID and Brook JD (1994). Holt-Oram syndrome is a genetically heterogeneous disease with one locus mapping to human chromosome 12q. *Nat Genet* **6**:401-404.

Thornton CA, Griggs RC and Moxley RT (1994). Myotonic dystrophy with no trinucleotide repeat expansion. *Annals of Neurol* **35**:269-272.

Timchenko L, Nastainczyk W, Schneider T, Hofmann F and Caskey CT (1995). Full-length myotonin protein kinase (72 kDa) displays serine kinase activity. *Proc Natl Acad Sci USA* **92**:5366-5370.

Töpert C, Döring F, Derst C, Daut J, Grzeschik K-H and Karschin A (2000). Cloning, structure and assignment to chromosome 19q13 of the human Kir2.4 inwardly rectifying potassium channel gene (*KCNJ14*). *Mam Genomics* **11**:247-249.

Torrington M, Weymar HW, Van der Merwe HW and Brink AJ (1986). Progressive familial heart block. Part I: Extent of the disease. *SA J Science* **70**:353-355.

Tóth G, Gáspári Z and Jurka J (2000). Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res* **10**:967-981.

The Utah Marker Development Group (1995). A collection of ordered tetranucleotide-repeat markers from the human genome. *Am J Hum Genet* **4**:1829-1836.

Vallianos G and Sideris DA (1974). Familial conduction defects. *Cardiol* **59:**190-197.

Van der Kooi AJ, van Meegen M, Ledderhof T, McNally EM, de Visser M and Bolhuis PA (1997). Genetic localisation of a newly synthesised autosomal dominant limb-girdle

muscular dystrophy with cardiac involvement (LGMD1B) to chromosome 1q11-21. *Am J Hum Genet* **60**:891-895.

Van der Merwe P-L, Weymar HW, Torrington M and Brink AJ (1986). Progressive familial heart block. Part II: Clinical and ECG confirmation of progression-report 4 cases. *S Afr Med J* **70**:356-357.

Van der Merwe P-L, Weymar HW, Torrington M and Brink A J (1988). Progressive familial heartblock (type I): a follow-up study after 10 years. *S Afr Med J* **73**:275-276.

Van der Merwe P-L, Rose AG, Van der Walt JJ, Weymer HW, Hunter JC and Weich HFH (1991). Progressive familial heart block, type I. *S Afr Med J* **70**:356-357.

Van der Merwe P-L, Weymar HW and Kalis NN (1994). The extent of progressive familial block type I: a new perspective. *SA J Cardiovasc Res* **5**:125-126.

Vatta M, Li H and Towbin JA (2000). Molecular biology of arrhythmic syndromes. *Curr Opin Cardiol* **15**:12-22.

Venter CJ, Adams MD, Myers EW, Li PW, Mural RW, Sutton GG, *et al.* (2001). The sequence of the human genome. *Science* **291**:1304-1359.

Vijg J and van Orsouw NJ (1999). Two-dimensional gene scanning: exploring human genetic variability. *Electrophoresis* **20**:1239-49.

Vitadello M, Matteoli M and Gorza L (1990). Neurofilament proteins are co-expressed with desmin in heart conduction system myocytes. *J Cell Science* **97**:11-21.

Vogel F and Motulsky AG (1986). *Human Genetics problems and approaches* (2nd ed). Springer Verlag, Berlin, Germany.

Vozzi C, Dupont E, Coppen SR, Yeh Hi and Severs NJ (1999). Chamber-related differences in connexin expression in the human heart. *J Mol Cell Biol* **31**:991-1003.

Wang Q, Splawski I, Atkinson D, Li Z, Robinson JL, Moss AJ, Towbin JA and Keating MT (1995). *SCN5A* mutations associated with an inherited cardiac arrhythmia, long QT syndrome. *Cell* **80**:805-811.

Wang DG, Fan J, Siao C, Berno P, Young R, Sapolsky G, *et al.* (1998). Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280**:1077-1082.

Warren ST (1996). The expanding world of trinucleotide repeats. Science 271:1374-1375.

Weber JL and May PE (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Hum Genet* **44**:388-396.

Weber JL (1990). Informativeness of human $(dC\text{-}dA)_n.(dG\text{-}dT)_n$ polymorphisms. *Genomics* **7**:524-530.

Weber JL, Wang Z, Hansen K, Stephenson M, Kappel C, Salzman S, *et al.* (1993). Evidence for human meiotic recombination interference obtained through of a short tandem repeat-polymorphism linkage map of chromosome 19. *Am J Hum Genet* **53**:1079-1095.

Weissenbach J, Gyapay G, Dib C, Vignal A, Morissette J, Millaseau P, *et al.* (1992). A second-generation linkage map of the human genome. *Nature* **359**:794-801.

Wolfsberg TG and Landsman D (1997). A comparison of expression sequence tags (ESTs) to human genomic sequences. *Nucleic Acids Res* **25**:1626-1623.

Zipes DP and Jalife J (2001). *Cardiac electrophysiology: from cell to bedside* (3rd ed). WB Saunders Co., USA.

APPENDIX I

SOLUTIONS, BUFFERS AND METHODS

1. STOCK SOLUTIONS

10x TBE BUFFER

Tris-OH (Merck)	0.89M, pH8.0
Boric acid (Merck)	0.89M
Na_2EDTA (Merck)	20mM, pH8.0

10x TE BUFFER STOCK SOLUTIONS

Tris-HCL (Merck)	0.1M, pH8.0
EDTA (Boehringer Mannheim)	0.01M, pH8.0

ETHIDIUM BROMIDE (10mg/ml)

Ethidium bromide (Sigma)	500mg
ddH_2O	50ml

Stir well on magnetic stirrer for 4 hours, store in a dark container at 4°C.

10% AMMONIUM PERSULPHATE

Ammonium persulphate (Merck)	5g
ddH_2O	50ml

Mix well and store at 4°C

2. DNA EXTRACTION

CELL LYSIS BUFFER

Sucrose (Merck)	0.32M
Triton X-100 (Merck)	1%
$MgCl_2$ (Merck)	5mM
Tris-HCl, pH 8.0 (Biorad)	10mM

ddH_2O to 1 litre

3M NaAc

NaAc.$3H_2O$ (Merck) 40.81g

Adjust to pH to 5.2 with glacial acetic acid (Merck) and adjust volume to 100ml with ddH_2O.

Na-EDTA SOLUTION

NaCl (Merck) 18.75ml of 4M stock solution

EDTA (Merck) 250ml of 100mM stock solution

Mix well.

PHENOL/CHLOROFORM

Mix phenol (Merck) (saturated with 1xTE), chloroform (Merck) and 8-hydroxyquinoline (Merck) in a ratio of 25:24:1. Store at 4°C.

CHLOROFORM/OCTANOL

Mix chloroform (Merck) and octanol (Merck) in a ratio of 24:1. Store at 4°C.

3. BACTERIAL CELL GROWTH MEDIA

LURIA-BERTANI MEDIUM

NaCl (Merck) 10g

Tryptone (Biolab) 10g

Yeast extract (Oxoid) 5g

ddH_2O to 1 litre

Adjust to pH7.2 with 5M NaOH and autoclave.

LURIA-BERTANI AGAR

NaCl 10g

Tryptone 10g

Yeast extract 5g

ddH_2O to 1 litre

Dissolve in 700ml ddH_2O and adjust to pH7.2 with 5M NaOH (Merck).

Bacto-agar (Difco Laboratories) 15g

ddH_2O to 1 litre and autoclave

4. COSMID DNA EXTRACTION

Using the QIAGEN^{R} plasmid/cosmid purification kit buffers (Stratagene). (See section 14).

5. PLASMID AND BAC DNA EXTRACTION

SOLUTION I (Resuspension buffer)

Glucose (BDH Chemicals)	50mM
Tris (pH8.0) (Merck)	25mM
EDTA (Boehringer Mannheim)	0.5M
ddH_2O to 500ml and autoclave	

SOLUTION II (Lysis buffer)

NaOH (Merck)	0.2M
SDS (Sigma)	1%
ddH_2O to 1ml	

SOLUTION III (Neutralising buffer)

Glacial acetic acid (Merk)	11.5%
Potassium acetate (Merck)	3M
ddH_2O to 100ml	

6. COLONY BLOT SOLUTIONS

DENATURING SOLUTIONS

NaCl (Merck)	1.5M
NaOH (Merck)	0.5M
ddH_2O to 1 litre	

NEUTRALISING SOLUTION

NaCl (Merck)	1.5M
Tris (pH 8.0) (Merck)	0.5M
ddH_2O to 1 litre	

PRE-WASH SOLUTION

SSC	5x
SDS (Sigma)	0.5%
EDTA (Boehringer Mannheim)	1mM
ddH_2O to 1 litre	

7. HYBRIDISATION STOCK SOLUTIONS

20xSSC

NaCl (Merck)	0.15M
Na citrate (BDH Chemicals)	0.15M

100X DENHARDT'S SOLUTION

Ficoll 400 (Merck)	0.025mM
Polyvinylpyrrolidine	0.027mM
BSA fraction	2% (g/ml)
Store at –20°C.	

20X SSPE

NaCl	125mM
Na_2HPO4 (BDH Chemicals)	17mM
NaH_2PO_4 (BDH Chemicals)	8mM
SDS (Sigma)	173mM
PEG_{8000} (Pharmacia)	12.5mM

SOLUTIONS FOR MEMBRANE FILTERS

PRE-HYBRIDISATION SOLUTION

20x SSPE	20ml
100X Denhardt's	4ml
10% SDS	1.6ml
Sonicated *E. coli* DNA (Sigma)	5mg/ml
ddH_2O to 80ml	

HYBRIDISATION WASH SOLUTION

SSC	2x
SDS	0.1%

8. AGAROSE GEL SOLUTIONS

0.7% AGAROSE GEL

Agarose (Seakem)	0.35g
10x TBE	5ml
H_2O	45ml

1% AGAROSE GEL

Agarose	0.5g
10x TBE	5ml
H_2O	45ml

1,5% AGAROSE GEL

Agarose	0.75g
10x TBE	5ml
H_2O	45ml

2% AGAROSE GEL

Agarose	1.0g
10x TBE	5ml
H_2O	45ml

Melt agarose (FMC) in a microwave oven on HIGH power, swirling occasionally until it is completely dissolved. Cool the solution to 50°C –60°C and add 5 µl ethidium bromide (10mg/ml) prior to casting.

9. POLYACRYLAMIDE GEL SOLUTIONS

12% NON-DENATURING POLYACRYLAMIDE GEL SOLUTION

40% acrylamide stock solution	8ml
10x TBE stock solution	2ml
ddH_2O	10ml
10% ammonium persulphate	160µl
TEMED	60µl

5% DENATURING POLYACRYLAMIDE GEL SOLUTION

40% acrylamide stock solution	10.8ml
10x TBE stock solution	6ml
Urea	26.64g
ddH_2O	12.75ml

Stir well with a magnetic stirrer until all urea is dissolved, then add

10% ammonium persulphate	500µl
TEMED	50µl

METHOD

CASTING OF SEQUENCING GELS

Glass plates were prepared by carefully and meticulously washing both plates with detergent. The plates were wiped with 70% ethanol, and were dried thoroughly to remove any traces of detergent. The notched plate was then silanised with Wynn's C-thru windshield rain dispersant (Wynn Oil S.A. Pty Ltd, SA) so that this plate could be removed while the gel remained fixed to the large plate, and the notched plate was placed silanised side-down on top of the spacers. This assembly was sealed with the use of a rubber boot (S2 casting boot, Life ™ Technologies) and the gel poured with the glass assembly at a slight angle. The glass assembly was placed horizontally and the loading front was formed with a top spacer of identical thickness to the lengthwise spacers. The gel was allowed to set for 1-2 hours, whereupon the top spacer was removed and replaced with a sharkstooth comb to separate wells for loading of samples.

10. SSCP ANALYSIS GEL SOLUTIONS

0.35x MDE GEL SOLUTION

2x MDE™ stock solution (FMC)	28ml
10x TBE stock solution	9.7ml
H_2O	122ml
10% ammonium persulphate	1.6ml
TEMED	160µl

5% NON-DENATURING POLYACRYLAMIDE GEL SOLUTION WITH 10% GLYCEROL

40% acrylamide stock solution (BDH Chemicals)	27ml
10x TBE stock solution	8ml
Glycerol (Merck)	16ml
H_2O	108ml
10% ammonium persulphate	1.6ml
TEMED	160µl

10% NON-DENATURING POLYACRYLAMIDE GEL SOLUTION WITH 10% GLYCEROL

40% acrylamide stock solution	54ml
10x TBE stock solution	8ml
Glycerol	16ml
H_2O	81ml
10% ammonium persulphate	1.6ml
TEMED (Promega)	160µl

8% MILDLY-DENATURING POLYACRYLAMIDE GEL SOLUTION

40% acrylamide stock solution	32.4ml
10x TBE stock solution	8ml
Urea	24g
Glycerol	8ml
H_2O	91.8ml
10% ammonium persulphate	1.6ml
TEMED	160µl

10% MILDLY-DENATURING POLYACRYLAMIDE GEL SOLUTION

40% acrylamide stock solution	40.5ml
10x TBE stock solution	8ml
Urea	24g
Glycerol	8ml
H_2O	84ml
10% ammonium persulphate	1.6ml
TEMED	160µl

METHOD

CASTING OF SSCP GELS

Glass plates were prepared by carefully and meticulously washing both plates with detergent. The plates were wiped with 70% ethanol, and were dried thoroughly to remove any traces of detergent. The notched plate was then silanised with Wynn's C-thru windshield rain dispersant (Wynn Oil S.A. Pty Ltd, SA) to ease the separation of the gel from the un-notched plate upon dismantling after electrophoresis. More 70% ethanol was sprayed on to the inner surface of the large glass plate. A sheet of Gelbond™ (FMC Bioproducts, Rockland, Maine, USA), cut to the same size as the large glass plate, was placed, hydrophobic side down, on this plate, taking care not to remove the protective sheet of paper from the hydrophilic side of the nylon membrane. Air bubbles trapped between the Gelbond™ and the glass were removed by careful rubbing over the protective sheet of paper with a piece of paper towel. Subsequently, this sheet of paper was removed from the nylon membrane and spacers of 1mm (SSCP analysis gels)

thickness placed lengthwise at the edges of the large glass plate. The notched plate was placed silanised side-down on top of these spacers. This assembly was subsequently sealed with the use of rubber boot (S2 rubber casting boot, Life™ Technologies) and the gel poured with the glass assembly slanted at a slight angle. The glass assembly was subsequently laid down horizontally and a square-tooth well-forming comb inserted. The gel was allowed to set for 1-2 hours.

11. LOADING DYES

BROMOPHENOL BLUE LOADING DYE FOR AGAROSE AND POLYACRYLAMIDE GEL ELECTROPHORESIS

Bromphenol blue (BDH Chemicals)	0.1% (w/v)
Glycerol (Merck)	60% (v/v)
Tris-Cl, pH8 (Merck)	10mM

All the components were mixed. The blue colour intensity of the loading dye was decreased compared to the conventional loading dye, in order to prevent masking of small PCR amplified DNA fragments on agarose gels.

SSCP LOADING DYE

Bromophenol blue (BDH Chemicals)	0.1% (w/v)
Xylene cyanol (M & B Laboratory Chemicals)	0.1% (w/v)
Formamide (Merck)	95%
NaOH	10mM
EDTA	20mM

Mix well. Aliquot and store at –20°C.

SEQUENCING STOP SOLUTION

Bromophenol blue (BDH Chemicals)	0.1% (w/v)
Xylene cyanol (M & B Laboratory Chemicals)	0.1% (w/v)
Formamide (Merck)	95% (v/v)

12. SILVER STAINING OF NON-DENATURING POLYACRYLAMIDE GELS

$AgNO_3$ SOLUTION (0.1%)

$AgNO_3$ (Merck)	1g
ddH_2O	1 litre

DEVELOPING SOLUTION

NaOH	15g
$NaBH_4$ (Sigma)	0.1g
Formaldehyde	4ml
ddH_2O to 1 litre	

Mix well on magnetic stirrer until all NaOH pellets are dissolved.

METHOD

SILVER STAINING OF POLYACRYLAMIDE GELS

The gel casting plates were dismantled and the polyacrylamide gel was gently immersed into a shallow receptacle containing freshly prepared $AgNO_3$ solution. The gel was agitated for 10 min, on a shaking platform. The $AgNO_3$ solution was poured off and the gel was thoroughly washed with ddH_2O. Freshly prepared developing solution was added and the gel was agitated in the solution until DNA fragments become visible as bands on the gel. The developing solution was poured off, the gel rinsed in ddH_2O, air-dried (SSCP gels) and photographed.

13. X-RAY FILM DEVELOPING

DEVELOPING SOLUTION

Polycon A variable contrast X-ray developer polycon A (Champion Photochemistry Pty Ltd, Midrand, SA) solution	1.98 litre
Polycon A variable contrast X-ray developer solution 2	18.1ml
ddH_2O	9.9 litre

STOP SOLUTION

Glacial acetic acid	540ml
Make up to 9 litres with ddH_2O	

FIXING SOLUTION

Perfix (Champion Photochemistry Pty Ltd, Midrand, SA)	1.8 litre
Hardener-S (Champion Photochemistry Pty Ltd, Midrand, SA)	225ml
Make up to 9 litres with H_2O.	

METHOD

The X-ray was clamped to a stainless steel frame under red light, and was immersed into the developing solution for 4 min, then into the stop solution for 30 sec and finally into

the fixer solution for 2 min. The film was rinsed well with H_2O, and then allowed to air dry.

14. METHODS FOR THE IDENTIFICATION OF RECOMBINANT COSMIDS CONTAINING $(A_3G)_n$ MOTIFS

Cosmid culture preparation and Colony blots

The recombinant Lawrist-16 cosmid cultures were prepared by dipping a sterile loop into a stab culture or by scraping the top of frozen glycerol stocks. The loop was then used to inoculate 2ml of Luria Bertani (LB) broth in 15ml tubes containing kanamycin (Kn) (50µg/ul). The tubes were incubated at 37°C overnight with vigorous shaking and subsequently 2ml of the culture were streaked onto a LB agar (LBA) plate containing 50µg/ml of Kn, which was incubated overnight, at 37°C. A single colony was picked from the incubated agar plate, inoculated into 250ml LB/Kn (50µg/ml) medium and incubated at 37°C overnight with vigorous shaking.

Isolation and purification of cosmid DNA

QIAGEN Method

The QIAGEN™ commercial DNA purification kit (QIAGEN, Germany) was used to isolate cosmid DNA from bacterial genomic DNA with slight modifications to the manufacturer's protocol. Cells were harvested from 250ml overnight culture by centrifugation at 4°C for 10 min at 6000rpm in a Sorvall RC-5B (Du Pont, USA) centrifuge using a GSA rotor. The supernatant was discarded and the pellet carefully and completely resuspended in 50ml of buffer P1 containing RNAse A (100µg/ml). The cells were then lysed by adding 50ml of buffer P2 to the suspension, which was then gently mixed and incubated for 5 min at room temperature. Fifty millilitres of chilled buffer P3 was added to the lysed cells and this was incubated on ice for 10 min. Bacterial DNA was precipitated from the insoluble cosmid DNA by centrifugation at 20000g (11000rpm) in a Sorvall RC-5B centrifuge with a GSA rotor for 30 min. The cosmid DNA-containing supernatant was transferred into a sterile GSA tube, re-centrifuged, for 15 min until the supernatant was clear. The supernatant was transferred into a clean tube. The next part of the purification involved gel filtration.

A QIAGEN-tip500 column was first equilibrated with 10ml of QBT buffer (QIAGEN, Germany) (Appendix I) before loading the cosmid containing supernatant. The next step involved washing the column twice with 30ml buffer QC to remove contaminants from the DNA. The DNA was released from the column by the addition of 15 ml of buffer QF and the eluant was collected in a SS34 tube. Cosmid DNA precipitation was performed by the addition of 10.5ml (0.7 volumes) of isopropanol at room temperature and centrifugation at 15000g (9500rpm) in a Sorvall RC-5B centrifuge using a SS34 rotor for 30 min at 4°C. The supernatant was carefully removed and the DNA-containing pellet washed with 5ml ethanol (70%), followed by centrifugation at 15000x g in a Sorvall RC-5B centrifuge using a SS34 rotor, for 10 min at 4°C. Ethanol was decanted and the remnants blotted with a paper towel. The pellet was air-dried for 5 min and resuspended

in 200µl 1xTE buffer (pH8.0). Absorbance was read on a spectrophotometer at 260nm to determine the DNA concentration (µg/ml) using the formula: OD_{260} x 50 x Dilution factor ($1OD_{260}$ = 50µg). The DNA was analysed on a 0.7% agarose gel containing 0.01% EtBr and 1xTBE to check its integrity (section 2.7.1). The cosmid DNA could now be utilised for the preparation of dot blots, Southern or sub-cloning

Preparation of Dot blots

Dot blots were used for the initial screening of cosmids to identify those cosmids containing $(A_3G)_n$ repeats and again later to verify the presence of $(A_3G)_n$ repeats in the inserts. The denaturation of the cosmid DNA was performed in the following steps: To each 200ng of cosmid DNA an equal volume of 0.4N NaOH was added and the mixture was incubated at room temperature for 5 min. The mixture was snap-cooled on ice for 2 min and spotted onto the previously marked positions on a Hybond^{TM} N^+ filter (Amersham, UK), which had been placed on 3MM-chromatography paper (Whatman, England).Fifty nanograms of human genomic DNA were spotted onto the filter as a positive control and ddH_2O as a negative control. The filter was air–dried for 15 min and thereafter baked for 2 hrs in an 80°C vacuum oven (Towson and Mecer Ltd, England). Subsequently, the filter was either re-hydrated in 2x SSC prior to the pre-hybridisation step of the screening process (section 2.10.1.6) or covered with clingfilm (GLADwrap^{TM}) and stored at 4°C until required.

Cosmid DNA digestion and Southern blot preparation

The digestion reaction, which contained 30µg cosmid DNA, 90U *Sau*3AI (Promega, USA), 10x buffer as specified by the manufacturer (Promega,USA), and ddH_2O to 150µl, was incubated at 37°C for 3 hrs (see fig. 2.5A). The digestion was monitored by electrophoresis of a 2µl aliquot of the reaction mix on a 0.7% agarose gel (Appendix I). When digestion was considered to be complete, the digested fragments were separated on a 20 x 20cm 0.7% agarose gel containing 0.01% EtBr (Appendix I) by electrophoresis at 35V overnight. In addition, λ *Pst*I, (Promega, USA) was run alongside the digested cosmid inserts as a marker. The following day a luminescent ruler was placed alongside the gel on an UV-transilluminator (3UVM^{TM} Transilluminator model, LMS-26E) and photographed using an ITC camera (Berkenhoff and Drebes, Germany) and a Sony video graphic printer (UP-860CE).

The gel containing the digested cosmid samples was prepared for Southern blotting in the following way. Immediately after the electrophoretic separation of the digested cosmid DNA fragments, the gel was depurinated by submersion and gentle agitation in 0.2M HCl for 10 min in a glass container. The HCl was decanted, and the gel rinsed three times in ddH_20 before placing it on a platform covered with 3MM Whatman paper for alkaline transfer of DNA fragments from the gel onto a Hybond^{TM} N^+ nucleic acid transfer membrane filter (Amersham, UK), following the manufacturer's protocol for alkaline Southern blotting (Amersham, UK). The DNA was fixed by baking the filter for 2 hrs at

80°C in a vacuum oven (Towson and Mercer Ltd, England). Thereafter, the Hybond^{TM} N^{+} membrane was either re-hydrated in 2x SSC (Appendix I) in preparation for probing with a radioactively labeled oligonucleotide or stored at 4°C.

14.5 Labeling of $(A_3G)_{10}$ oligonucleotides

In order to identify the $(A_3G)_n$ repeat containing fragments to be used for subcloning, the Hybond^{TM} N^{+} membrane was probed with an end labeled γ–^{32}P[dATP]-$(A_3G)_{10}$ oligonucleotide. The end-labeling reaction mixture contained 1µl of $(A_3G)_{10}$ oligonucleotide (50pmol), 10µl γ–^{32}P[dATP] (Amersham, UK), 2µl T4 polynucleotide kinase (20 units) (Promega, USA), 5µl of 10xT4 polynucleotide kinase buffer (Promega, USA), and 32µl ddH_2O. The labeling mixture was incubated for 2 hrs at 37°C.

Gel filtration chromatography on Sephadex G-25 was used to separate the radio-labeled probe from the unincorporated γ–^{32}P[dATP]. The incorporated radioactivity in the labeled oligonucleotide had to be at least $1x10^6$ cpm/ml of hybridisation solution before hybridisation was pursued.

14.6 Pre-hybridisation, hybridisation and post-hybridisation

The filters (Hybond^{TM} N^{+} membranes, dot blots or the colony blots) were prehybridised in twenty millilitres of pre-warmed (42°C) pre-hybridisation liquid (Appendix I), in a heat-sealable plastic bag, in a Perspex container, in a shaking waterbath (Heto Lab equipment, Denmark) at 42°C for 6 hrs. Thereafter, one corner of the bag was cut open and the pre-hybridisation fluid discarded. Twenty millilitres of fresh hybridisation solution (Appendix I) containing the radioactivity labeled probe was added to the bag which was then re-sealed and incubated overnight at 42°C, in a shaking water-bath (Heto Lab equipment, Denmark). After incubation, the hybridisation solution was discarded appropriately, and the filters washed. This was done by submerging the labeled filters in a flat-bottomed plastic box containing 2x SSC with 0.1% SDS (Appendix I), after which the filters were incubated for 30 min at room temperature with gentle shaking. This washing step was repeated 4 times, with fresh 2x SSC with 0.1% SDS for each wash. During the washes the filters were checked with a Geiger counter during and after each wash, so that the washing step was stopped when the radioactivity dropped to 10 counts per sec since most of the background radioactivity would have been removed. The filters were dried at room temperature on a sheet of 3MM-chromatography paper (Whatman), sealed in a plastic bag. The filters were exposed to an X-ray film (Cronex-4, Protea Medical) for 48 hrs and were developed as described in section 2.7.4. In the case of filters, positive signals observed on the autoradiograph represented inserts containing $(A_3G)_n$ repeats. The method used to identify and isolate these inserts is described in section 14.7. The identified $(A_3G)_n$ repeat containing DNA was investigated further as described in section 2.10.2.

14.7 Insert DNA preparation

*Sau*3A1 restriction enzyme digestion of pBluescript SK DNA

For the purposes of cloning, 30µg of cosmid DNA containing the $(A_3G)_n$ repeat were digested with 90U of *Sau*3A1 (Amersham, UK) according to the manufacturer's protocol for 3 hrs and electrophoresed on 0.7% agarose gel (20cm x 20cm x 3cm) (Appendix I) at 30V overnight alongside λ *Pst*1 (Promega, USA). The following day, the gel was photographed and the fragments containing the $(A_3G)_n$ repeat were identified by comparison of the positive signals from the autoradiograph (section 2.10.1.6) and the λ *Pst*1 size marker as a guide. These fragments were excised from the gel using a sterile blade and then the DNA was extracted from the gel slice using a Talent kit (Cleanmix, Italy) according to the manufacturer's protocol (as described in section 2.6.3.1) and resuspended in 40µl ddH_2O.

Partial fill-in reactions for insert

In order to prevent self-ligation of the insert molecules and to render the ends compatible to the partially filled in vector, dGTP and dATP were added to the insert in the presence of the Klenow fragment of DNA polymerase (Promega, USA). This was then referred to as a partially filled-in insert. The fill-in reaction mixture consisted of 40µl (10ng) of insert DNA, 0.5µl dGTP (10mM), 0.5µl dATP (10mM), 2µl (10U) Klenow fragment of DNA polymerase (Promega, USA), 5µl 10x Klenow buffer (Promega, USA) and ddH_2O made up to 50µl. Incubation was carried out at 37°C for 2 hrs, after which 1µl (0.5M) EDTA was added to terminate each reaction and the reaction tube was stored at 4°C.

14.8 Vector preparation

***Sal*I restriction enzyme digestion of pBluescript SK DNA**

Plasmid vector pBluescript SK (Stratagene, USA) (fig. 2.4) cultures were prepared in a similar manner to the cosmid, as described in section 2.10.1.1 except LB/ampicillin (Amp) (50µg/ml) selective medium was used. The plasmid vector, pBluescript SK, molecular weight 2.9kb, was linearised by digesting with *Sal*1 restriction enzyme (Promega, USA). The digestion reaction mixture contained 10µg vector DNA, 4µl (100mM) spermidine, 200U *Sal*I (Promega, USA), 10µl 10x buffer as specified by the manufacturer (Promega, USA) and ddH_2O to a final volume of 100µl. The reaction was incubated at 37°C for 3 hrs. To verify that the vector had been linearised, 5µl of digestion reaction mixture was analysed on a 0.7% agarose gel (section 2.7.1), with an equal volume of an uncut vector sample and λ *Pst*1 size marker in adjacent lanes. Thereafter, linearised vector DNA in the remaining sample was precipitated by the addition of 250µl 100% ethanol. The mixture was vortexed briefly and incubated for 30 min at –80°C, prior to centrifugation in a microfuge (model Microfuge® Lite, Beckmann Instruments, Germany) at 14000rpm for 15 min at room temperature. The resulting supernatant was carefully decanted and the pellet washed twice with 200µl 70% ethanol, followed by centrifugation in a microfuge (model Microfuge® Lite, Beckmann, Germany) at 14000rpm for 10 min. The ethanol was decanted, and the pellet lyophilised in a Speedy-

Vac concentrator (Savant Instruments, USA) for 5 min and then resuspended in 40µl ddH_2O. As the 5'-overhanging ends generated by *Sal*1 digestion of plasmid vector and the *Sau*3A1 digestion of insert DNA were incompatible for ligation, partial fill-in reactions were first performed for both insert and vector DNA.

Partial fill-in of vector

The phospate group at the 5'-end of a linearised vector has a tendency to self ligate in the presence of T4 DNA ligase (Damak and Bullock 1993). For this reason, the ends of the vectors were rendered non-complementary by the addition of deoxy-thymidine triphosphate (dTTP) and deoxy-cytosine (dCTP) to the linearised vector in the presence of the Klenow fragment of DNA polymerase (Promega, USA) then referred to as a partially filled vector. The fill-in reaction mixture consisted of 40µl of vector DNA, 0.5µl dTTP (10mM), 0.5µl dCTP (10mM), 2µl (10U) Klenow fragment of DNA polymerase (Promega, USA), 5µl 10x Klenow buffer and ddH_2O to 50µl. Incubation was carried out at 37°C for 2 hrs, after which 1µl (0.5M) EDTA was added to quench each reaction, which were stored at 4°C.

Self ligation of vector

The fill-in reaction of the vector may not be completely successful for all the vector molecules, with the result that a proportion of the vector will not yield transformants, which are non-recombinants. Therefore, to increase the efficiency of cloning, the partially filled-in vector was self-ligated and the resulting circularised vector was separated from the linearised vector by electrophoresis, based on their different migration distances in agarose. The self ligation reaction contained 5µl 10x ligation buffer (Appendix I), 5µl of 5mM ATP, 5µl 10mM of dithiothreitol (DTT), 34µl partially filled-in vector and 5U T4 ligase (Boehringer Mannheim, Germany) and incubation was performed overnight at room temperature or at 4°C. The following day, the ligation mixture was electrophoresed on a 0.7% agarose gel (Appendix I) next to λ *Pst*I marker, 25ng of linearised vector and 2µl uncut pBluescript SK (Stratagene, USA) to distinguish the linearised from the circularised vector and visualised under UV light. The linear vector was cut out with a sterile blade and placed in a sterile microfuge tube. The DNA was extracted using the Talent kit (Cleanmix, Italy) according to the manufacturer's protocol and stored at 4°C.

14.9 Subcloning of $(A_3G)_n$ repeat fragments

The cosmid vector is approximately 8kb in size and can house an insert of up to 42kb in size; since Taq polymerase can generate fragments for sequencing up to 1kb in length, the insert was digested into smaller fragments of about 1kb in size. These fragments were then subcloned into pBluescript SK (Stratagene, USA), from which sequencing was initiated by using standard vector primers.

Organic extraction of filled insert and vector

For subcloning, the partially filled insert and vector were purified by organic extraction. The partial fill-in reactions of both vector and insert were purified by the same method. To each tube 50µl TE, 1.5µl 5M NaCl and 100µl of phenol/chloroform (Appendix I) were added. The reagents were agitated by vortexing and centrifuged in a microfuge (model Microfuge® Lite, Beckmann, Germany) at 14000rpm for 5 min. The DNA-containing (top) aqueous phase was transferred into a clean 1.5ml eppendorf tube and after a 100µl aliquot of chloroform/isoamylalcohol was added, the sample was vortexed for 15 sec and then centrifuged as before. The top layer was carefully transferred to a clean 1.5ml eppendorf tube without disturbing the interface, after which 300µl of 100% ethanol was added and the sample incubated at –70°C for 15 min to precipitate the DNA. The DNA was pelleted by centrifugation at 14000rpm for 15 min at 4°C in a microfuge (model Microfuge® Lite, Beckmann, Germany). Both insert DNA and vector DNA-containing pellets were washed in 70% ethanol, lyophilised in a Speedy-Vac concentrator (Savant Instruments, USA) for 5 min and resuspended in 40µl ddH_2O.

Ligation

The partial fill-in reactions created compatible cohesive termini for both insert and vector to facilitate the ligation process, which resulted in the formation of a chimeric molecule. In order to find the optimal vector:insert ratios, 1µl of insert at varying dilutions of 1/10, 1/50 and 1/100 in H_2O were used, while keeping the volume of vector constant (table 1, ligation numbers 1, 2, 3 and 4). The one control reaction, ligation number 5, excluded the use of insert DNA, in order to determine the number of false recombinants formed as a result of the vector re-ligating to itself, i.e. establishing the efficiency of the partial fill-in reaction. The second control was the exclusion of vector DNA, to assess the competency of the *E. coli* XL1 Blue strain (table 2.5, ligation number 6). The ligation reactions were incubated at 16°C overnight in an Endocal refrigerated circulating bath (Neslab, Massachusetts, USA) and thereafter, stored at 4°C until required for transformation.

Table 1 Ligation reactions of insert DNA to pBluescript vector.

Ligation number	Vector µl	Insert µl	ATP 5mM	DTT 10mM	Ligation Buffer µl	T4 ligase 3U/µl	H_2O µl
1	1	1	1	1	1	1	4
2	1	1 (1/10)	1	1	1	1	4
3	1	1 (1/50)	1	1	1	1	4
4	1	1 (1/100)	1	1	1	1	4
5	1	0	1	1	1	1	5
6	0	0	1	1	1	1	6

Transformation

E. coli XL1-blue competent cells (Stratagene, USA) were prepared using the calcium chloride protocol (Sambrook *et al.*, 1989). Five microlitres of ligation mixture (table 1) were added to pre-chilled 15ml falcon tubes (Centrolab, Italy), followed by the addition of a 50ul aliquot of thawed competent cells and the resulting mixture was incubated on ice for 30 min. Meanwhile, in another tube, equal volumes (800µl) of 5-bromo-4-chloro-3-indoyl-β-D-galactopyranoside (X-gal) (20mg/ml) and isopropyl-β-D-thio-galactopyranoside (IPTG) (20mg/ml) were mixed together, of which 80µl were spread onto a LBA plate containing Amp (50µg/ml) and Tetracycline (Tet) (15µg/ml). After incubation, the chilled competent cells were heat-shocked by incubation in a water-bath at 42°C for 45 sec and were then placed on ice for 1 min, after which 500µl of LB at 37°C was added to the transformed cells, which were then left in a shaking incubator at 37°C for 1 to 2 hrs. Subsequently, 200µl of transformed cells were spread on the LBA plates containing IPTG and X-gal using 10-15 autoclaved beads (2-3mm in diameter) and the plates were incubated inverted at 37°C overnight. An aliquot of uncut pBR322 was added to the competent cells, to confirm the efficiency of transformation.

Colony lifts of recombinant of clones

Colony lifts were performed to identify recombinant clones containing the $(A_3G)_n$ repeat. Plates containing recombinant clones were placed at 4°C for at least 1 hr after overnight incubation at 37°C and prior to the colony procedure. For each master plate, two pieces of Hybond N^+ (Amersham, UK) membrane were cut to the size of the petri plate and marked with a pencil as master or duplicate, for replica plating. Using forceps, the first or master filter, was carefully placed on the plate with colonies for 1 minute and holes keyed in with an 18 gauge needle, taking care not to make any lateral movements, to act as a reference for the orientation of the filter on the plate. The master filter was carefully removed and placed, colony side up, on a fresh LBA/Amp/Tc plate. A duplicate filter was made by carefully placing the second filter on top of the master filter with the holes keyed in the same position as the master filter. The fresh plate with sandwiched duplicate filters and the original plate were incubated for 5 hrs at 37°C, after which the original plate was covered with with clingfilm (GLADwrap™) and stored at 4°C stored until required.

Following incubation, the filter sandwich was removed from the plate with the two halves kept together. Thereafter, the filters were separated, each one placed colony side up and they were treated sequentially with 10% SDS (Sigma, USA) for 3 min, denaturing solution (Appendix I) for 5 min and then with neutralising solution (Appendix I) for 5 min. This was performed in three separate containers layered with 3MM blotting paper (Whatman, England), which had been previously soaked, but not submerged in these solutions. Each filter was then transferred to a container with 2x SSC-soaked (Appendix I) 3MM blotting paper (Whatman, England) for a 5 min wash, before being air-dried for 30 min on dry 3MM blotting paper (Whatman, England). The filters were baked for 2 hrs in an 80°C vacuum oven (Towson and Mercer Ltd, England). Thereafter, the filters were

either covered with clingfilm (GLADwrap^{TM}) and stored at 4°C, or re-hydrated in 2x SSC for 5 min in preparation for probing with radioactively labeled oligonucleotide (section 2.10.1.5). The filters were finally washed and exposed to X-ray film overnight at room temperature. After development of the film, duplicate positive colonies were identified and selected from the original plate for sequencing.

Identification and characterisation of $(A_3G)_n$ repeat motif

To verify that the positive colonies picked from the plates contained cloned insert, recombinant plasmid DNA was extracted (section 2.10.1.2) and digested with *Kpn*I and *Sac*I at the vector restriction enzyme sites on either side of the multiple cloning site (MCS) (fig. 2.6). This digestion reaction contained 5µl plasmid DNA, 3-5U *Kpn*I (Promega, USA), 3-5U *Sac*I (Promega, USA), 1µl 10x buffer according to the manufacturer's (Promega, USA) specifications and ddH_2O to 10µl. The reaction mixture was incubated at 37°C for 2 hrs. Thereafter, 5µl of sample was withdrawn and 2µl loading dye added to it, after which the sample was loaded onto a 0.7% agarose gel (section 2.7.1) for analysis of digest reaction products. The DNA size marker λ *Pst*I was electrophoresed alongside the sample, as well as an undigested recombinant plasmid. Following the verification of the presence of an insert, the corresponding recombinant plasmid clones were sequenced (section 2.6.3) using vector primers T3 and T7 (fig. 2.4).

APPENDIX II

WEBSITES

Database/Programme	Description and electronic address
BCM searchlauncher	An integrated website which allows access to a number of bioinformatics tools, Baylor College of Medicine, Houston, Texas,USA http://www.hgsc.bcm.tcm.edu
CLUSTALW 1.81	Multiple DNA sequence alignment programme distributed by the European Bioinformatics Institute (EBI) at Cambrige, UK http://www.ebi.ac.uk/clustalw
Ensembl	Ensembl is a joint project between European Molecular Biology Laboratories (EMBL), EBI and the Sanger Institute, which develop a software system producing and maintaining automatic annotation on metazoan genomes and is funded by the Wellcome Trust. http://www.ensemb.org/
EST2GENOME	DNA alignment programme http://ftp.sanger.ac.uk/pub/EMBOSS
Genbank	DNA and protein sequence databases distributed by the USA National Centre for Biotechnology (NCBI), NIH http://www.ncbi.nlm.nih.gov
BLAST	A Basic Local Alignment Search Tool (BLAST) for gene and protein similarity searches http://www.ncbi.nlm.nih.gov/BLAST
BLASTp	A Basic Local Alignment Search Tool (BLAST) for gene and protein similarity searches http://www.ncbi.nlm.nih.gov/BLAST

EST	Collection of expressed sequence tags (ESTs) http://www.ncbi.nlm.nih.gov/dbEST
htgs	High throughput genomic sequence database http://www.ncbi.nlm.gov/htgs
human genome	Human genome DNA sequence database http://www.ncbi.nlm.nih.gov/human genome
nr	Nonredundant DNA sequence database http://www.ncbi.nlm.gov/nr
nucleotide	collection of nucleotide sequences (http://www.ncbi.nlm.nih/nucleotide).
SNP	Collection of single nucleotide polymorphisms (SNPs) http://www.ncbi.nih.gov/SNP
STS	Collection of sequence tag sites (STSs) http://www.ncbi.nih.nlm.gov/STS
Genemap99	Mapping information of human genomes http://www.ncbi.nlm.nih.gov/Genemap99
OMIM	Online Mendelian Inheritance in Man (OMIM) contains information on human genes and genetic disorders http://www.ncbi.nih.gov/OMIM
Locuslink	Provides descriptive information about genetic loci http://ncbi.nih.nlm.gov/LOCUSLINK
UNIGENE	Short gene transcript clustering programme (htpp://www.ncbi.nlm.nih.gov/UNIGENE).
Genomescan	Identifies the functional domains of proteins. http://genomescan.gov were used to identify
Genoscope	The French National Sequencing Center facility, which has a database of human sequences http//www.genoscope.org
JGI	Joint Genome Institute, which houses chromosomal sequence data and was formed by the merging of Los Alamos National Laboratory and the Lawrence Berkeley National Laboratory

	http://www.jgi.doe.gov
	JGI ftp site collaborative containing chromosome 19 sequence (ftp://sawdoff.llnl.gov/pub/JGI_data/Human/Ch19)
LLNL	A US Department of Energy national laboratory operated by the University of California involved in the establishment of the human chromosome 19 sequencing database http//:www-bio.llnl.gov/genome/
Primer	PCR primer design programme http://eatworms.swmed.edu/~tim/primerfinder
Repeatmasker	Programme used to remove *E. coli* sequences from human draft sequence (http://ftp.genome.washington.edu/cgi-bin/RepeatMasker
Stanford Human Genome Centre	Completion of the sequencing of chromosome 19 http://www-shgc.stanford.edu
STACKpack	Gene transcript clustering programme. Developed at South African National Institute of Bioinformatics (SANBI), SA. http://www.sanbi.ac.za/STACK
SWISS-PROT	Protein database. Maintained at University of Geneva Access via NCBI and EBI htpp://www.ncbi.nlm.nih.gov/ or http://www.ebi.ac.uk
Tandem Repeatfinder	A programme designed to detect long stretches of tandem repeats (http://c3biomath.mssm.edu/trf.basic.submit.html)

APPENDIX III

THE CLUSTAL ALIGNMENTS OF SCREENED CANDIDATE GENESFOR MUTATIONS (SEE SECTION 3.5 OF RESULTS)

The arrangement of the clustal alignment of sequences of DNA samples was presented as follows: **Sample 1** is the database sequence of the different exons with the flanking intronic sequence. The exonic sequence is indicated in **red** and the flanking intronic sequence is indicated in **black**. **Sample 2** is the sequence from the coding strand from an unaffected individual that does not have the disease-associated haplotype. **Sample 3** is the reverse complement of the reverse sequence of sample 2. **Sample 4** is the sequence from the coding strand from an affected PFHBI individual of pedigree 1, carrying the disease-associated haplotype. **Sample 5** is the reverse complement of the reverse sequence of sample 4. **Sample 6** is the sequence from the coding strand from an affected PFHBI individual of pedigree 2, carrying the disease-associated haplotype. **Sample 7** is the reverse complement of the reverse sequence of sample 6. **Sample 8** is the sequence from the coding strand from an affected PFHBI individual of pedigree 5, carrying the disease-associated haplotype. **Sample 9** is the reverse complement of the reverse sequence of sample 8. The alignment results for each exon with its flanking sequence of *KIR2.4* (tables 1-6), *LIN-7B* (tables 7-11), *BAX* (tables 12-16) and *GSY1* (tables 17-33) are presented.

A summary of **exons** in which sequence changes were detected is presented in table 3.14. and the legends in the following tables are shown in **bold**.

The sequencing data of *KCNA7* is not presented in this section but can be found in section 3.6.1.3 of results.

KIR2.4

Table 1 CLUSTALW 1.81 alignment of exon 1 of *KIR2.4*

```
1                                   GAAGATCATCACCCCCTTTCACAGTTGGGGAACAGGCTCAGGGAACG 47
2                                   GAAGATCATCACCCCCTTTCACAGTTGGGGAACAGGCTCAGGGAACG
4                                   GAAGATCATCACCCCCTTTCACAGTTGGGGAACAGGCTCAGGGAACG
5                                   GAAGATCATCACCCCCTTTCACAGTTGGGGAACAGGCTCAGGGAACG
6                                   GAAGATCATCACCCCCTTTCACAGTTGGGGAACAGGCTCAGGGAACG
7                                   GAAGATCATCACCCCCTTTCACAGTTGGGGAACAGGCTCAGGGAACG
9                                   GAAGATCATCACCCCCTTTCACAGTTGGGGAACAGGCTCAGGGAACG
                                    **************************************************

1                   CCATGAGCTGGCCAGGCCCCCGCAGTGCCAGGGGGAATGCTGGTTGGGTCCTAAGCCTTT 107
2                   CCATGAGCTGGCCAGGCCCCCGCAGTGCCAGGGGGAATGCTGGTTGGGTCCTAAGCCTTT
4                   CCATGAGCTGGCCAGGCCCCCGCAGTGCCAGGGGGAATGCTGGTTGGGTCCTAAGCCTTT
5                   CCATGAGCTGGCCAGGCCCCCGCAGTGCCAGGGGGAATGCTGGTTGGGTCCTAAGCCTTT
6                   CCATGAGCTGGCCAGGCCCCCGCAGTGCCAGGGGGAATGCTGGTTGGGTCCTAAGCCTTT
7                   CCATGAGCTGGCCAGGCCCCCGCAGTGCCAGGGGGAATGCTGGTTGGGTCCTAAGCCTTT
9                   CCATGAGCTGGCCAGGCCCCCGCAGTGCCAGGGGGAATGCTGGTTGGGTCCTAAGCCTTT
                    ************************************************************

1                   CTTTCCTTCCCAGAGTTCTGCCCTCGAAATCCTCCAGTCCCACCTCAGCCCCGTGAGGCA 167
2                   CTTTCCTTCCCAGAGTTCTGCCCTCGAAATCCTCCAGTCCCACCTCAGCCCCGTGAGGCA
4                   CTTTCCTTCCCAGAGTTCTGCCCTCGAAATCCTCCAGTCCCACCTCAGCCCCGTGAGGCA
5                   CTTTCCTTCCCAGAGTTCTGCCCTCGAAATCCTCCAGTCCCACCTCAGCCCCGTGAGGCA
6                   CTTTCCTTCCCAGAGTTCTGCCCTCGAAATCCTCCAGTCCCACCTCAGCCCCGTGAGGCA
7                   CTTTCCTTCCCAGAGTTCTGCCCTCGAAATCCTCCAGTCCCACCTCAGCCCCGTGAGGCA
9                   CTTTCCTTCCCAGAGTTCTGCCCTCGAAATCCTCCAGTCCCACCTCAGCCCCGTGAGGCA 2
                    ************************************************************

1                   GACCCGTTGTCTACTTCCCAAAGCCAACGGGAACCTTCACCAGGGGGCCATCTGCGGCCA 227
2                   GACCCGTTGTCTACTTCCCAAAGCCAACGGGAACCTTCACCAGGGGGCCATCTGCGGCCA
4                   GACCCGTTGTCTACTTCCCAAAGCCAACGGGAACCTTCACCAGGGGGCCATCTGCGGCCA
5                   GACCCGTTGTCTACTTCCCAAAGCCAACGGGAACCTTCACCAGGGGGCCATCTGCGGCCA
6                   GACCCGTTGTCTACTTCCCAAAGCCAACGGGAACCTTCACCAGGGGGCCATCTGCGGCCA
7                   GACCCGTTGTCTACTTCCCAAAGCCAACGGGAACCTTCACCAGGGGGCCATCTGCGGCCA
9                   GACCCGTTGTCTACTTCCCAAAGCCAACGGGAACCTTCACCAGGGGGCCATCTGCGGCCA
                    ************************************************************

1                   ACATGGCCCAGCTCCCAGGCTGGCTGCGACTCAGGCTCAGGGCACCCCCCTCCCCTAGCA 287
2                   ACATGGCCCAGCTCCCAGGCTGGCTGCGACTCAGGCTCAGGGCACCCCCCTCCCCTAGCA
4                   ACATGGCCCAGCTCCCAGGCTGGCTGCGACTCAGGCTCAGGGCACCCCCCTCCCCTAGCA
5                   ACATGGCCCAGCTCCCAGGCTGGCTGCGACTCAGGCTCAGGGCACCCCCCTCCCCTAGCA
6                   ACATGGCCCAGCTCCCAGGCTGGCTGCGACTCAGGCTCAGGGCACCCCCCTCCCCTAGCA
7                   ACATGGCCCAGCTCCCAGGCTGGCTGCGACTCAGGCTCAGGGCACCCCCCTCCCCTAGCA
9                   ACATGGCCCAGCTCCCAGGCTGGCTGCGACTCAGGCTCAGGGCACCCCCCTCCCCTAGCA
                    ************************************************************

1.                  AAAGCAGGGCGGGAGGGAAAGGGAT 312
2                   AAAGCAGGGCGGGAGGGAAAGGGAT
4                   AAAGCAGGGCGGGAGGGAAAGGGAT
5                   AAAGCAGGGCGGGAGGGAAAGGGAT
6                   AAAGCAGGGCGGGAGGGAAAGGGAT
7                   AAAGCAGGGCGGGAGGGAAAGGGAT
9                   AAAGCAGGGCGGGAGGGAAAGGGAT
                    *************************
```

Exonic sequence of the database sequence is shown in red (from position 25 to 172) and is 148bp in size. The size of the PCR product analysed is 312bp, which includes the exonic and the flanking intronic sequence.

1 = sequence of the exon from the database (http://www.ncbi.nlm.nih/nucleotide)

2 = forward sequence
3 = reverse complement sequence of the reverse } an unaffected individual

4 = forward sequence
5 = reverse complement sequence of the reverse } affected individual from pedigree 1

6 = forward sequence
7 = reverse complement sequence of the reverse } affected individual from pedigree 2

8 = forward sequence
9 = reverse complement sequence of the reverse } affected individual from pedigree 5

Table 2 CLUSTALW 1.81 alignment of exon 2A of *KCNJ14*

```
1                                                       CACGGAAGTCCCGCCTCTCCTCCCGG 28
2                                                       CACGGAAGTCCCGCCTCTCCTCCCGG
3                                                       CACGGAAGTCCCGCCTCTCCTCCCGG
4                                                       CACGGAAGTCCCGCCTCTCCTCCCGG
5                                                       CACGGAAGTCCCGCCTCTCCTCCCGG
6                                                       CACGGAAGTCCCGCCTCTCCTCCCGG
7                                                       CACGGAAGTCCCGCCTCTCCTCCCGG
8                                                       CACGGAAGTCCCGCCTCTCCTCCCGG
9                                                       CACGGAAGTCCCGCCTCTCCTCCCGG
                                                        **************************

1               GCGCACCTGCAGCAGCTGGGCACGCACGTGGGCCTCGACCAGGTGGCTGCGGCGCAGGTT 88
2               GCGCACCTGCAGCAGCTGGGCACGCACGTGGGCCTCGACCAGGTGGCTGCGGCGCAGGTT
3               GCGCACCTGCAGCAGCTGGGCACGCACGTGGGCCTCGACCAGGTGGCTGCGGCGCAGGTT
4               GCGCACCTGCAGCAGCTGGGCACGCACGTGGGCCTCGACCAGGTGGCTGCGGCGCAGGTT
5               GCGCACCTGCAGCAGCTGGGCACGCACGTGGGCCTCGACCAGGTGGCTGCGGCGCAGGTT
6               GCGCACCTGCAGCAGCTGGGCACGCACGTGGGCCTCGACCAGGTGGCTGCGGCGCAGGTT
7               GCGCACCTGCAGCAGCTGGGCACGCACGTGGGCCTCGACCAGGTGGCTGCGGCGCAGGTT
8               GCGCACCTGCAGCAGCTGGGCACGCACGTGGGCCTCGACCAGGTGGCTGCGGCGCAGGTT
9               GCGCACCTGCAGCAGCTGGGCACGCACGTGGGCCTCGACCAGGTGGCTGCGGCGCAGGTT
                ************************************************************

1               GCCGACGCGCCACATGAGGCAGAGGCGGTGGTCGCGCAGCGCCACGACGGCGTTCTCGCT 148
2               GCCGACGCGCCACATGAGGCAGAGGCGGTGGTCGCGCAGCGCCACGACGGCGTTCTCGCT
3               GCCGACGCGCCACATGAGGCAGAGGCGGTGGTCGCGCAGCGCCACGACGGCGTTCTCGCT
4               GCCGACGCGCCACATGAGGCAGAGGCGGTGGTCGCGCAGCGCCACGACGGCGTTCTCGCT
5               GCCGACGCGCCACATGAGGCAGAGGCGGTGGTCGCGCAGCGCCACGACGGCGTTCTCGCT
6               GCCGACGCGCCACATGAGGCAGAGGCGGTGGTCGCGCAGCGCCACGACGGCGTTCTCGCT
7               GCCGACGCGCCACATGAGGCAGAGGCGGTGGTCGCGCAGCGCCACGACGGCGTTCTCGCT
8               GCCGACGCGCCACATGAGGCAGAGGCGGTGGTCGCGCAGCGCCACGACGGCGTTCTCGCT
9               GCCGACGCGCCACATGAGGCAGAGGCGGTGGTCGCGCAGCGCCACGACGGCGTTCTCGCT
                ************************************************************

1               GAAGACCAGCGTCTCGTTGCGCTTCTTGGGTTTGGCCATCTTGGCCATGACAGCACCCAC 208
2               GAAGACCAGCGTCTCGTTGCGCTTCTTGGGTTTGGCCATCTTGGCCATGACAGCACCCAC
3               GAAGACCAGCGTCTCGTTGCGCTTCTTGGGTTTGGCCATCTTGGCCATGACAGCACCCAC
4               GAAGACCAGCGTCTCGTTGCGCTTCTTGGGTTTGGCCATCTTGGCCATGACAGCACCCAC
5               GAAGACCAGCGTCTCGTTGCGCTTCTTGGGTTTGGCCATCTTGGCCATGACAGCACCCAC
6               GAAGACCAGCGTCTCGTTGCGCTTCTTGGGTTTGGCCATCTTGGCCATGACAGCACCCAC
7               GAAGACCAGCGTCTCGTTGCGCTTCTTGGGTTTGGCCATCTTGGCCATGACAGCACCCAC
8               GAAGACCAGCGTCTCGTTGCGCTTCTTGGGTTTGGCCATCTTGGCCATGACAGCACCCAC
9               GAAGACCAGCGTCTCGTTGCGCTTCTTGGGTTTGGCCATCTTGGCCATGACAGCACCCAC
                ************************************************************

1               GACGAAGGCGTCGAGCACGCAGCCGGCAATGCACTGCAGCACCACGGCGGCCACAGCGGC 268
2               GACGAAGGCGTCGAGCACGCAGCCGGCAATGCACTGCAGCACCACGGCGGCCACAGCGGC
3               GACGAAGGCGTCGAGCACGCAGCCGGCAATGCACTGCAGCACCACGGCGGCCACAGCGGC
4               GACGAAGGCGTCGAGCACGCAGCCGGCAATGCACTGCAGCACCACGGCGGCCACAGCGGC
5               GACGAAGGCGTCGAGCACGCAGCCGGCAATGCACTGCAGCACCACGGCGGCCACAGCGGC
6               GACGAAGGCGTCGAGCACGCAGCCGGCAATGCACTGCAGCACCACGGCGGCCACAGCGGC
7               GACGAAGGCGTCGAGCACGCAGCCGGCAATGCACTGCAGCACCACGGCGGCCACAGCGGC
8               GACGAAGGCGTCGAGCACGCAGCCGGCAATGCACTGCAGCACCACGGCGGCCACAGCGGC
9               GACGAAGGCGTCGAGCACGCAGCCGGCAATGCACTGCAGCACCACGGCGGCCACAGCGGC
                ************************************************************
```

```
1          CGGGCACTCCTCGGTGACGCTGCGCACGCCGTAGCCGATGGACGTCTGCGTCTCCAGCGC 328
2          CGGGCACTCCTCGGTGACGCTGCGCACGCCGTAGCCGATGGACGTCTGCGTCTCCAGCGC
3          CGGGCACTCCTCGGTGACGCTGCGCACGCCGTAGCCGATGGACGTCTGCGTCTCCAGCGC
4          CGGGCACTCCTCGGTGACGCTGCGCACGCCGTAGCCGATGGACGTCTGCGTCTCCAGCGC
5          CGGGCACTCCTCGGTGACGCTGCGCACGCCGTAGCCGATGGACGTCTGCGTCTCCAGCGC
6          CGGGCACTCCTCGGTGACGCTGCGCACGCCGTAGCCGATGGACGTCTGCGTCTCCAGCGC
7          CGGGCACTCCTCGGTGACGCTGCGCACGCCGTAGCCGATGGACGTCTGCGTCTCCAGCGC
8          CGGGCACTCCTCGGTGACGCTGCGCACGCCGTAGCCGATGGACGTCTGCGTCTCCAGCGC
9          CGGGCACTCCTCGGTGACGCTGCGCACGCCGTAGCCGATGGACGTCTGCGTCTCCAGCGC
           ************************************************************

1          GAAGAGGAAGGCGGCCAGGAAGCTGGCCACGTGTGAGAAGCAGGGCGCGGGCGGTGGCGG 368
2          GAAGAGGAAGGCGGCCAGGAAGCTGGCCACGTGTGAGAAGCAGGGCGCGGGCGGTGGCGG
3          GAAGAGGAAGGCGGCCAGGAAGCTGGCCACGTGTGAGAAGCAGGGCGCGGGCGGTGGCGG
4          GAAGAGGAAGGCGGCCAGGAAGCTGGCCACGTGTGAGAAGCAGGGCGCGGGCGGTGGCGG
5          GAAGAGGAAGGCGGCCAGGAAGCTGGCCACGTGTGAGAAGCAGGGCGCGGGCGGTGGCGG
6          GAAGAGGAAGGCGGCCAGGAAGCTGGCCACGTGTGAGAAGCAGGGCGCGGGCGGTGGCGG
7          GAAGAGGAAGGCGGCCAGGAAGCTGGCCACGTGTGAGAAGCAGGGCGCGGGCGGTGGCGG
8          GAAGAGGAAGGCGGCCAGGAAGCTGGCCACGTGTGAGAAGCAGGGCGCGGGCGGTGGCGG
9          GAAGAGGAAGGCGGCCAGGAAGCTGGCCACGTGTGAGAAGCAGGGCGCGGGCGGTGGCGG
           ************************************************************

1          GGCGGCCAGGT 379
2          GGCGGCCAGGT
3          GGCGGCCAGGT
4          GGCGGCCAGGT
5          GGCGGCCAGTT
6          GGCGGCCAGGT
7          GGCGGCCAGGT
8          GGCGGCCAGGT
9          GGCGGCCAGGT
           ***********
```

Exon 2 (740bp) was subdivided into two parts, namely exon 2A (371bp) and 2B (369bp). The database exonic sequence of 2A (from position 9 to 379) is shown in red. The overlapping exonic sequence of 2A with 2B is underlined. The size of the PCR product analysed is shown (379bp), which includes the flanking intronic sequence, exonic sequence of 2A and the flanking exonic sequence of 2B.

1 = sequence of the exon from the database (http://www.ncbi.nlm.nih/nucleotide)

2 = forward sequence 3 = reverse complement sequence of the reverse	an unaffected individual
4 = forward sequence 5 = reverse complement sequence of the reverse	affected individual from pedigree 1
6 = forward sequence 7= reverse complement sequence of the reverse	affected individual from pedigree 2
8 = forward sequence 9 = reverse complement sequence of the reverse	affected individual from pedigree 5

Table 3 CLUSTALW 1.81 alignment of exon 2B of *Kir2.4*

```
1                                                                  GGCGGGGCGGCCAGGTC 380
2                                                                  GGCGGGGCGGCCAGGTC
3                                                                  GGCGGGGCGGCCAGGTC
4                                                                  GGCGGGGCGGCCAGGTC
5                                                                  GGCGGGGCGGCCAGGTC
6                                                                  GGCGGGGCGGCCAGGTC
7                                                                  GGCGGGGCGGCCAGGTC
8                                                                  GGCGGGGCGGCCAGGTC
9                                                                  GGCGGGGCGGCCAGGTC
                                                                   *****************

1                  GCCGTGCAGCGAGGCAATGAGCCAGAAGGCCAGGCCGAAGAGCAGCCAGGAGGCGAGGAA 440
2                  GCCGTGCAGCGAGGCAATGAGCCAGAAGGCCAGGCCGAAGAGCAGCCAGGAGGCGAGGAA
3                  GCCGTGCAGCGAGGCAATGAGCCAGAAGGCCAGGCCGAAGAGCAGCCAGGAGGCGAGGAA
4                  GCCGTGCAGCGAGGCAATGAGCCAGAAGGCCAGGCCGAAGAGCAGCCAGGAGGCGAGGAA
5                  GCCGTGCAGCGAGGCAATGAGCCAGAAGGCCAGGCCGAAGAGCAGCCAGGAGGCGAGGAA
6                  GCCGTGCAGCGAGGCAATGAGCCAGAAGGCCAGGCCGAAGAGCAGCCAGGAGGCGAGGAA
7                  GCCGTGCAGCGAGGCAATGAGCCAGAAGGCCAGGCCGAAGAGCAGCCAGGAGGCGAGGAA
8                  GCCGTGCAGCGAGGCAATGAGCCAGAAGGCCAGGCCGAAGAGCAGCCAGGAGGCGAGGAA
9                  GCCGTGCAGCGAGGCAATGAGCCAGAAGGCCAGGCCGAAGAGCAGCCAGGAGGCGAGGAA
                   ************************************************************

1                  GGAGCAGGAGAAGAGCAGGCACATCCAGCGCCAGCGCACGTCCACGCATGTGGTGAACAG 500
2                  GGAGCAGGAGAAGAGCAGGCACATCCAGCGCCAGCGCACGTCCACGCATGTGGTGAACAG
3                  GGAGCAGGAGAAGAGCAGGCACATCCAGCGCCAGCGCACGTCCACGCATGTGGTGAACAG
4                  GGAGCAGGAGAAGAGCAGGCACATCCAGCGCCAGCGCACGTCCACGCATGTGGTGAACAG
5                  GGAGCAGGAGAAGAGCAGGCACATCCAGCGCCAGCGCACGTCCACGCATGTGGTGAACAG
6                  GGAGCAGGAGAAGAGCAGGCACATCCAGCGCCAGCGCACGTCCACGCATGTGGTGAACAG
7                  GGAGCAGGAGAAGAGCAGGCACATCCAGCGCCAGCGCACGTCCACGCATGTGGTGAACAG
8                  GGAGCAGGAGAAGAGCAGGCACATCCAGCGCCAGCGCACGTCCACGCATGTGGTGAACAG
9                  GGAGCAGGAGAAGAGCAGGCACATCCAGCGCCAGCGCACGTCCACGCATGTGGTGAACAG
                   ************************************************************

1                  GTCGCTCAGGTAGCGCGCGCCCTGGCCACCCAGGTTTACGAAACGCACGTTGCAGTGCCC 560
2                  GTCGCTCAGGTAGCGCGCGCCCTGGCCACCCAGGTTTACGAAACGCACGTTGCAGTGCCC
3                  GTCGCTCAGGTAGCGCGCGCCCTGGCCACCCAGGTTTACGAAACGCACGTTGCAGTGCCC
4                  GTCGCTCAGGTAGCGCGCGCCCTGGCCACCCAGGTTTACGAAACGCACGTTGCAGTGCCC
5                  GTCGCTCAGGTAGCGCGCGCCCTGGCCACCCAGGTTTACGAAACGCACGTTGCAGTGCCC
6                  GTCGCTCAGGTAGCGCGCGCCCTGGCCACCCAGGTTTACGAAACGCACGTTGCAGTGCCC
7                  GTCGCTCAGGTAGCGCGCGCCCTGGCCACCCAGGTTTACGAAACGCACGTTGCAGTGCCC
8                  GTCGCTCAGGTAGCGCGCGCCCTGGCCACCCAGGTTTACGAAACGCACGTTGCAGTGCCC
9                  GTCGCTCAGGTAGCGCGCGCCCTGGCCACCCAGGTTTACGAAACGCACGTTGCAGTGCCC
                   ************************************************************

1                  GTCTTTCTTGACGAAGCGACCGCGGCGCCGGCCCACGGGTGACTGCACCGGTGCCGGCGC 620
2                  GTCTTTCTTGACGAAGCGACCGCGGCGCCGGCCCACGGGTGACTGCACCGGTGCCGGCGC
3                  GTCTTTCTTGACGAAGCGACCGCGGCGCCGGCCCACGGGTGACTGCACCGGTGCCGGCGC
4                  GTCTTTCTTGACGAAGCGACCGCGGCGCCGGCCCACGGGTGACTGCACCGGTGCCGGCGC
5                  GTCTTTCTTGACGAAGCGACCGCGGCGCCGGCCCACGGGTGACTGCACCGGTGCCGGCGC
6                  GTCTTTCTTGACGAAGCGACCGCGGCGCCGGCCCACGGGTGACTGCACCGGTGCCGGCGC
7                  GTCTTTCTTGACGAAGCGACCGCGGCGCCGGCCCACGGGTGACTGCACCGGTGCCGGCGC
8                  GTCTTTCTTGACGAAGCGACCGCGGCGCCGGCCCACGGGTGACTGCACCGGTGCCGGCGC
9                  GTCTTTCTTGACGAAGCGACCGCGGCGCCGGCCCACGGGTGACTGCACCGGTGCCGGCGC
                   ************************************************************
```

```
1  CCACCCGTTGCGGCACAACCCGGGCCCGGCCTCCTCTTCATCGCCCGCCCGGCTGTCTCC 680
2  CCACCCGTTGCGGCACAACCCGGGCCCGGCCTCCTCTTCATCGCCCGCCCGGCTGTCTCC
3  CCACCCGTTGCGGCACAACCCGGGCCCGGCCTCCTCTTCATCGCCCGCCCGGCTGTCTCC
4  CCACCCGTTGCGGCACAACCCGGGCCCGGCCTCCTCTTCATCGCCCGCCCGGCTGTCTCC
5  CCACCCGTTGCGGCACAACCCGGGCCCGGCCTCCTCTTCATCGCCCGCCCGGCTGTCTCC
6  CCACCCGTTGCGGCACAACCCGGGCCCGGCCTCCTCTTCATCGCCCGCCCGGCTGTCTCC
7  CCACCCGTTGCGGCACAACCCGGGCCCGGCCTCCTCTTCATCGCCCGCCCGGCTGTCTCC
8  CCACCCGTTGCGGCACAACCCGGGCCCGGCCTCCTCTTCATCGCCCGCCCGGCTGTCTCC
9  CCACCCGTTGCGGCACAACCCGGGCCCGGCCTCCTCTTCATCGCCCGCCCGGCTGTCTCC
   ************************************************************

1  CGAATCCAGGGCGCCGCTGAGGCGGCGTAGGGCCCTGGCCAGGCCCATTGGACGGAGGGA 740
2  CGAATCCAGGGCGCCGCTGAGGCGGCGTAGGGCCCTGGCCAGGCCCATTGGACGGAGGGA
3  CGAATCCAGGGCGCCGCTGAGGCGGCGTAGGGCCCTGGCCAGGCCCATTGGACGGAGGGA
4  CGAATCCAGGGCGCCGCTGAGGCGGCGTAGGGCCCTGGCCAGGCCCATTGGACGGAGGGA
5  CGAATCCAGGGCGCCGCTGAGGCGGCGTAGGGCCCTGGCCAGGCCCATTGGACGGAGGGA
6  CGAATCCAGGGCGCCGCTGAGGCGGCGTAGGGCCCTGGCCAGGCCCATTGGACGGAGGGA
7  CGAATCCAGGGCGCCGCTGAGGCGGCGTAGGGCCCTGGCCAGGCCCATTGGACGGAGGGA
8  CGAATCCAGGGCGCCGCTGAGGCGGCGTAGGGCCCTGGCCAGGCCCATTGGACGGAGGGA
9  CGAATCCAGGGCGCCGCTGAGGCGGCGTAGGGCCCTGGCCAGGCCCATTGGACGGAGGGA
   ************************************************************

1  CCCCCTCCACTTGGCCTAGTGGGGGGCAGGCGCCCCCAAC 800
2  CCCCCTCCACTTGGCCTAGTGGGGGGCAGGCGCCCCCAAC
3  CCCCCTCCACTTGGCCTAGTGGGGGGCAGGCGCCCCCAAC
4  CCCCCTCCACTTGGCCTAGTGGGGGGCAGGCGCCCCCAAC
5  CCCCCTCCACTTGGCCTAGTGGGGGGCAGGCGCCCCCAAC
6  CCCCCTCCACTTGGCCTAGTGGGGGGCAGGCGCCCCCAAC
7  CCCCCTCCACTTGGCCTAGTGGGGGGCAGGCGCCCCCAAC
8  CCCCCTCCACTTGGCCTAGTGGGGGGCAGGCGCCCCCAAC
9  CCCCCTCCACTTGGCCTAGTGGGGGGCAGGCGCCCCCAAC
   ****************************************
```

The database exonic sequence is shown in red (from position 364 to 758bp). The size of the PCR product analysed is shown (438bp), which includes the flanking exonic sequence from 2A, exonic sequence of 2B and the flanking intronic sequence. The overlapping exonic sequence 2A with 2B is underlined.

1 = sequence of the exon from the database (http://www.ncbi.nlm.nih/nucleotide)

2 = forward sequence 3 = reverse complement sequence of the reverse	an unaffected individual
4 = forward sequence 5 = reverse complement sequence of the reverse	affected individual from pedigree 1
6 = forward sequence 7 = reverse complement sequence of the reverse	affected individual from pedigree 2
8 = forward sequence 9 = reverse complement sequence of the reverse	affected individual from pedigree 5

Table 4 CLUSTALW 1.81 alignment of exon 3A of *KIR2.4*

```
1                                                                  CAGCCCTGGGCCTCC 15
2                                                                  CAGCCCTGGGCCTCC
3                                                                  CAGCCCTGGGCCTCC
4                                                                  CAGCCCTGGGCCTCC
5                                                                  CAGCCCTGGGCCTCC
6                                                                  CAGCCCTGGGCCTCC
7                                                                  CAGCCCTGGGCCTCC
8                                                                  CAGCCCTGGGCCTCC
                                                                   ***************

1                CCCTTCCCTTGTCCACTAATTAGCCCTGTGCCCTACAGGGTCTCACCTTCACCTTCAC 75
2                CCCTTCCCTTGTCCACTAATTAGCCCTGTGCCCTACAGGGTCTCACCTTCACCTTCAC
3                CCCTTCCCTTGTCCACTAATTAGCCCTGTGCCCTACAGGGTCTCACCTTCACCTTCAC
4                CCCTTCCCTTGTCCACTAATTAGCCCTGTGCCCTACAGGGTCTCACCTTCACCTTCAC
5                CCCTTCCCTTGTCCACTAATTAGCCCTGTGCCCTACAGGGTCTCACCTTCACCTTCAC
6                CCCTTCCCTTGTCCACTAATTAGCCCTGTGCCCTACAGGGTCTCACCTTCACCTTCAC
7                CCCTTCCCTTGTCCACTAATTAGCCCTGTGCCCTACAGGGTCTCACCTTCACCTTCAC
8                CCCTTCCCTTGTCCACTAATTAGCCCTGTGCCCTACAGGGTCTCACCTTCACCTTCAC
                 ************************************************************

1                CTTCCAGGTCAGGCCAGAAATGTTGCTTCTGAAGTCCTGAGTCCACCTCCTAGCACTGCC 135
2                CTTCCAGGTCAGGCCAGAAATGTTGCTTCTGAAGTCCTGAGTCCACCTCCTAGCACTGCC
3                CTTCCAGGTCAGGCCAGAAATGTTGCTTCTGAAGTCCTGAGTCCACCTCCTAGCACTGCC
4                CTTCCAGGTCAGGCCAGAAATGTTGCTTCTGAAGTCCTGAGTCCACCTCCTAGCACTGCC
5                CTTCCAGGTCAGGCCAGAAATGTTGCTTCTGAAGTCCTGAGTCCACCTCCTAGCACTGCC
6                CTTCCAGGTCAGGCCAGAAATGTTGCTTCTGAAGTCCTGAGTCCACCTCCTAGCACTGCC
7                CTTCCAGGTCAGGCCAGAAATGTTGCTTCTGAAGTCCTGAGTCCACCTCCTAGCACTGCC
8                CTTCCAGGTCAGGCCAGAAATGTTGCTTCTGAAGTCCTGAGTCCACCTCCTAGCACTGCC
                 ************************************************************

1                AAGCAGTCCACCTGGAGGTGATGAGAACTTCCAGGCTTTGGTAATGGACAACTCTGTATC 195
2                AAGCAGTCCACCTGGAGGTGATGAGAACTTCCAGGCTTTGGTAATGGACAACTCTGTATC
3                AAGCAGTCCACCTGGAGGTGATGAGAACTTCCAGGCTTTGGTAATGGACAACTCTGTATC
4                AAGCAGTCCACCTGGAGGTGATGAGAACTTCCAGGCTTTGGTAATGGACAACTCTGTATC
5                AAGCAGTCCACCTGGAGGTGATGAGAACTTCCAGGCTTTGGTAATGGACAACTCTGTATC
6                AAGCAGTCCACCTGGAGGTGATGAGAACTTCCAGGCTTTGGTAATGGACAACTCTGTATC
7                AAGCAGTCCACCTGGAGGTGATGAGAACTTCCAGGCTTTGGTAATGGACAACTCTGTATC
8                AAGCAGTCCACCTGGAGGTGATGAGAACTTCCAGGCTTTGGTAATGGACAACTCTGTATC
                 ************************************************************

1                CCTGGAGCCTTCATTCCTTTTGGCATTACAGAACCACAAGAAGTCATCTAGAATCATTGA 255
2                CCTGGAGCCTTCATTCCTTTTGGCATTACAGAACCACAAGAAGTCATCTAGAATCATTGA
3                CCTGGAGCCTTCATTCCTTTTGGCATTACAGAACCACAAGAAGTCATCTAGAATCATTGA
4                CCTGGAGCCTTCATTCCTTTTGGCATTACAGAACCACAAGAAGTCATCTAGAATCATTGA
5                CCTGGAGCCTTCATTCCTTTTGGCATTACAGAACCACAAGAAGTCATCTAGAATCATTGA
6                CCTGGAGCCTTCATTCCTTTTGGCATTACAGAACCACAAGAAGTCATCTAGAATCATTGA
7                CCTGGAGCCTTCATTCCTTTTGGCATTACAGAACCACAAGAAGTCATCTAGAATCATTGA
8                CCTGGAGCCTTCATTCCTTTTGGCATTACAGAACCACAAGAAGTCATCTAGAATCATTGA
                 ************************************************************

1                GATTCTCAACTCCATTGGGCAAACTGGAACCAGCAAAATCTCAGATATCTAGAATTGCCA 315
2                GATTCTCAACTCCATTGGGCAAACTGGAACCAGCAAAATCTCAGATATCTAGAATTGCCA
3                GATTCTCAACTCCATTGGGCAAACTGGAACCAGCAAAATCTCAGATATCTAGAATTGCCA
4                GATTCTCAACTCCATTGGGCAAACTGGAACCAGCAAAATCTCAGATATCTAGAATTGCCA
5                GATTCTCAACTCCATTGGGCAAACTGGAACCAGCAAAATCTCAGATATCTAGAATTGCCA
6                GATTCTCAACTCCATTGGGCAAACTGGAACCAGCAAAATCTCAGATATCTAGAATTGCCA
7                GATTCTCAACTCCATTGGGCAAACTGGAACCAGCAAAATCTCAGATATCTAGAATTGCCA
8                GATTCTCAACTCCATTGGGCAAACTGGAACCAGCAAAATCTCAGATATCTAGAATTGCCA
                 ************************************************************
```

```
1          CTGCTTCCTTGACTCTGTGATCAGTCTACAGCCACAGAATCCTGATTCCCTTTGGACATT 375
2          CTGCTTCCTTGACTCTGTGATCAGTCTACAGCCACAGAATCCTGATTCCCTTTGGACATT
3          CTGCTTCCTTGACTCTGTGATCAGTCTACAGCCACAGAATCCTGATTCCCTTTGGACATT
4          CTGCTTCCTTGACTCTGTGATCAGTCTACAGCCACAGAATCCTGATTCCCTTTGGACATT
5          CTGCTTCCTTGACTCTGTGATCAGTCTACAGCCACAGAATCCTGATTCCCTTTGGACATT
6          CTGCTTCCTTGACTCTGTGATCAGTCTACAGCCACAGAATCCTGATTCCCTTTGGACATT
7          CTGCTTCCTTGACTCTGTGATCAGTCTACAGCCACAGAATCCTGATTCCCTTTGGACATT
8          CTGCTTCCTTGACTCTGTGATCAGTCTACAGCCACAGAATCCTGATTCCCTTTGGACATT
           ************************************************************

1          CTAGAATGGCACTTTAAAAAAAAACAACCCACCCAGTGGATCCTTTCAAAGCACTGCTTA 435
2          CTAGAATGGCACTTTAAAAAAAAACAACCCACCCAGTGGATCCTTTCAAAGCACTGCTTA
3          CTAGAATGGCACTTTAAAAAAAAACAACCCACCCAGTGGATCCTTTCAAAGCACTGCTTA
4          CTAGAATGGCACTTTAAAAAAAAACAACCCACCCAGTGGATCCTTTCAAAGCACTGCTTA
5          CTAGAATGGCACTTTAAAAAAAAACAACCCACCCAGTGGATCCTTTCAAAGCACTGCTTA
6          CTAGAATGGCACTTTAAAAAAAAACAACCCACCCAGTGGATCCTTTCAAAGCACTGCTTA
7          CTAGAATGGCACTTTAAAAAAAAACAACCCACCCAGTGGATCCTTTCAAAGCACTGCTTA
8          CTAGAATGGCACTTTAAAAAAAAACAACCCACCCAGTGGATCCTTTCAAAGCACTGCTTA
           ************************************************************

1          TTCTCTTGGCTGGGTGCTCTACAGCCAGAAGTTGGAGTTATCTAATATCAAGAACTCTTA 495
2          TTCTCTTGGCTGGGTGCTCTACAGCCAGAAGTTGGAGTTATCTAATATCAAGAACTCTTA
3          TTCTCTTGGCTGGGTGCTCTACAGCCAGAAGTTGGAGTTATCTAATATCAAGAACTCTTA
4          TTCTCTTGGCTGGGTGCTCTACAGCCAGAAGTTGGAGTTATCTAATATCAAGAACTCTTA
5          TTCTCTTGGCTGGGTGCTCTACAGCCAGAAGTTGGAGTTATCTAATATCAAGAACTCTTA
6          TTCTCTTGGCTGGGTGCTCTACAGCCAGAAGTTGGAGTTATCTAATATCAAGAACTCTTA
7          TTCTCTTGGCTGGGTGCTCTACAGCCAGAAGTTGGAGTTATCTAATATCAAGAACTCTTA
8          TTCTCTTGGCTGGGTGCTCTACAGCCAGAAGTTGGAGTTATCTAATATCAAGAACTCTTA
           ************************************************************

1          ATCTGTCACTTCAGTCTTTCTAAACCCACAGACTAACAATGTTTTGTGCATTCTAGAAAC 555
2          ATCTGTCACTTCAGTCTTTCTAAACCCACAGACTAACAATGTTTTGTGCATTCTAGAAAC
3          ATCTGTCACTTCAGTCTTTCTAAACCCACAGACTAACAATGTTTTGTGCATTCTAGAAAC
4          ATCTGTCACTTCAGTCTTTCTAAACCCACAGACTAACAATGTTTTGTGCATTCTAGAAAC
5          ATCTGTCACTTCAGTCTTTCTAAACCCACAGACTAACAATGTTTTGTGCATTCTAGAAAC
6          ATCTGTCACTTCAGTCTTTCTAAACCCACAGACTAACAATGTTTTGTGCATTCTAGAAAC
7          ATCTGTCACTTCAGTCTTTCTAAACCCACAGACTAACAATGTTTTGTGCATTCTAGAAAC
8          ATCTGTCACTTCAGTCTTTCTAAACCCACAGACTAACAATGTTTTGTGCATTCTAGAAAC
           ************************************************************

1          ATTACTCTTTAGTTTCTGGAGTAGTTCAACTCTCCACGTAAAATCTGTCCTAGACTCACC 615
2          ATTACTCTTTAGTTTCTGGAGTAGTTCAACTCTCCACGTAAAATCTGTCCTAGACTCACC
3          ATTACTCTTTAGTTTCTGGAGTAGTTCAACTCTCCACGTAAAATCTGTCCTAGACTCACC
4          ATTACTCTTTAGTTTCTGGAGTAGTTCAACTCTCCACGTAAAATCTGTCCTAGACTCACC
5          ATTACTCTTTAGTTTCTGGAGTAGTTCAACTCTCCACGTAAAATCTGTCCTAGACTCACC
6          ATTACTCTTTAGTTTCTGGAGTAGTTCAACTCTCCACGTAAAATCTGTCCTAGACTCACC
7          ATTACTCTTTAGTTTCTGGAGTAGTTCAACTCTCCACGTAAAATCTGTCCTAGACTCACC
8          ATTACTCTTTAGTTTCTGGAGTAGTTCAACTCTCCACGTAAAATCTGTCCTAGACTCACC
           ************************************************************

1          AGTGCTTGCCTCTGTGGT 633
2          AGTGCTTGCCTCTGTGGT
3          AGTGCTTGCCTCTGTGGT
4          AGTGCTTGCCTCTGTGGT
5          AGTGCTTGCCTCTGTGGT
6          AGTGCTTGCCTCTGTGGT
7          AGTGCTTGCCTCTGTGGT
8          AGTGCTTGCCTCTGTGGT
           ******************
```

Exon 3 was subdivided into exons 3A, 3B and 3C (1789bp in size). The database exonic sequence 3A is shown in red (from position 4 to 633) and is 630bp in size. The size of the PCR product analysed is 633bp, which includes the flanking intronic sequence, exon 3A and part of exon 3B. The overlapping exonic sequence of 3A and 3B is underlined.

1 = sequence of the exon from the database (http://www.ncbi.nlm.nih/nucleotide)

2 = forward sequence
3 = reverse complement sequence of the reverse } an unaffected individual

4 = forward sequence
5 = reverse complement sequence of the reverse } affected individual from pedigree 1

6 = forward sequence
7 = reverse complement sequence of the reverse } affected individual from pedigree 2

8 = forward sequence
9 = reverse complement sequence of the reverse } affected individual from pedigree 5

Table 5 CLUSTALW 1.81 alignment of exon 3B of *KIR2.4*

```
1          GGTCTTTGGTCAATCAAGATGATCTTGAATACTTGGGGCATTCCATAACCTTCCCTCCTT 690
2          GGTCTTTGGTCAATCAAGATGATCTTGAATACTTGGGGCATTCCATAACCTTCCCTCCTT
3          GGTCTTTGGTCAATCAAGATGATCTTGAATACTTGGGGCATTCCATAACCTTCCCTCCTT
4          GGTCTTTGGTCAATCAAGATGATCTTGAATACTTGGGGCATTCCATAACCTTCCCTCCTT
5          GGTCTTTGGTCAATCAAGATGATCTTGAATACTTGGGGCATTCCATAACCTTCCCTCCTT
6          GGTCTTTGGTCAATCAAGATGATCTTGAATACTTGGGGCATTCCATAACCTTCCCTCCTT
7          GGTCTTTGGTCAATCAAGATGATCTTGAATACTTGGGGCATTCCATAACCTTCCCTCCTT
8          GGTCTTTGGTCAATCAAGATGATCTTGAATACTTGGGGCATTCCATAACCTTCCCTCCTT
9          GGTCTTTGGTCAATCAAGATGATCTTGAATACTTGGGGCATTCCATAACCTTCCCTCCTT
           ************************************************************

1          AGGTGGTCTAGAACAAATGATTTGCTACTCAGTTGGCTGAGTTGAGAAGCAACACTCCCT 750
2          AGGTGGTCTAGAACAAATGATTTGCTACTCAGTTGGCTGAGTTGAGAAGCAACACTCCCT
3          AGGTGGTCTAGAACAAATGATTTGCTACTCAGTTGGCTGAGTTGAGAAGCAACACTCCCT
4          AGGTGGTCTAGAACAAATGATTTGCTACTCAGTTGGCTGAGTTGAGAAGCAACACTCCCT
5          AGGTGGTCTAGAACAAATGATTTGCTACTCAGTTGGCTGAGTTGAGAAGCAACACTCCCT
6          AGGTGGTCTAGAACAAATGATTTGCTACTCAGTTGGCTGAGTTGAGAAGCAACACTCCCT
7          AGGTGGTCTAGAACAAATGATTTGCTACTCAGTTGGCTGAGTTGAGAAGCAACACTCCCT
8          AGGTGGTCTAGAACAAATGATTTGCTACTCAGTTGGCTGAGTTGAGAAGCAACACTCCCT
9          AGGTGGTCTAGAACAAATGATTTGCTACTCAGTTGGCTGAGTTGAGAAGCAACACTCCCT
           ************************************************************

1          TGAATCTTAGGTGAATCCAGACCTACAAACTCTTGATATCTGATGCTGTTCAAGAACTGC 810
2          TGAATCTTAGGTGAATCCAGACCTACAAACTCTTGATATCTGATGCTGTTCAAGAACTGC
3          TGAATCTTAGGTGAATCCAGACCTACAAACTCTTGATATCTGATGCTGTTCAAGAACTGC
4          TGAATCTTAGGTGAATCCAGACCTACAAACTCTTGATATCTGATGCTGTTCAAGAACTGC
5          TGAATCTTAGGTGAATCCAGACCTACAAACTCTTGATATCTGATGCTGTTCAAGAACTGC
6          TGAATCTTAGGTGAATCCAGACCTACAAACTCTTGATATCTGATGCTGTTCAAGAACTGC
7          TGAATCTTAGGTGAATCCAGACCTACAAACTCTTGATATCTGATGCTGTTCAAGAACTGC
8          TGAATCTTAGGTGAATCCAGACCTACAAACTCTTGATATCTGATGCTGTTCAAGAACTGC
9          TGAATCTTAGGTGAATCCAGACCTACAAACTCTTGATATCTGATGCTGTTCAAGAACTGC
           ************************************************************

1          CAACTTGCTTCCCCAGAGGCAGTCTATACCATTGGCTTCTCACCCTGGTCAACAATCCAG 870
2          CAACTTGCTTCCCCAGAGGCAGTCTATACCATTGGCTTCTCACCCTGGTCAACAATCCAG
3          CAACTTGCTTCCCCAGAGGCAGTCTATACCATTGGCTTCTCACCCTGGTCAACAATCCAG
4          CAACTTGCTTCCCCAGAGGCAGTCTATACCATTGGCTTCTCACCCTGGTCAACAATCCAG
5          CAACTTGCTTCCCCAGAGGCAGTCTATACCATTGGCTTCTCACCCTGGTCAACAATCCAG
6          CAACTTGCTTCCCCAGAGGCAGTCTATACCATTGGCTTCTCACCCTGGTCAACAATCCAG
7          CAACTTGCTTCCCCAGAGGCAGTCTATACCATTGGCTTCTCACCCTGGTCAACAATCCAG
8          CAACTTGCTTCCCCAGAGGCAGTCTATACCATTGGCTTCTCACCCTGGTCAACAATCCAG
9          CAACTTGCTTCCCCAGAGGCAGTCTATACCATTGGCTTCTCACCCTGGTCAACAATCCAG
           ************************************************************

1          GTCTCCCCATATAGAGGAATCTAGAACTGCCAGTCTTTTCCTCTTGGGTGATCCAGAATT 930
2          GTCTCCCCATATAGAGGAATCTAGAACTGCCAGTCTTTTCCTCTTGGGTGATCCAGAATT
3          GTCTCCCCATATAGAGGAATCTAGAACTGCCAGTCTTTTCCTCTTGGGTGATCCAGAATT
4          GTCTCCCCATATAGAGGAATCTAGAACTGCCAGTCTTTTCCTCTTGGGTGATCCAGAATT
5          GTCTCCCCATATAGAGGAATCTAGAACTGCCAGTCTTTTCCTCTTGGGTGATCCAGAATT
6          GTCTCCCCATATAGAGGAATCTAGAACTGCCAGTCTTTTCCTCTTGGGTGATCCAGAATT
7          GTCTCCCCATATAGAGGAATCTAGAACTGCCAGTCTTTTCCTCTTGGGTGATCCAGAATT
8          GTCTCCCCATATAGAGGAATCTAGAACTGCCAGTCTTTTCCTCTTGGGTGATCCAGAATT
9          GTCTCCCCATATAGAGGAATCTAGAACTGCCAGTCTTTTCCTCTTGGGTGATCCAGAATT
           ************************************************************
```

```
1          CTGGTGAGACTGAGGAATCAGGCTCATAAAGGGTTCTCAGGAGGGAGCACAGAAGCAGAG 990
2          CTGGTGAGACTGAGGAATCAGGCTCATAAAGGGTTCTCAGGAGGGAGCACAGAAGCAGAG
3          CTGGTGAGACTGAGGAATCAGGCTCATAAAGGGTTCTCAGGAGGGAGCACAGAAGCAGAG
4          CTGGTGAGACTGAGGAATCAGGCTCATAAAGGGTTCTCAGGAGGGAGCACAGAAGCAGAG
5          CTGGTGAGACTGAGGAATCAGGCTCATAAAGGGTTCTCAGGAGGGAGCACAGAAGCAGAG
           CTGGTGAGACTGAGGAATCAGGCTCATAAAGGGTTCTCAGGAGGGAGCACAGAAGCAGAG
7          CTGGTGAGACTGAGGAATCAGGCTCATAAAGGGTTCTCAGGAGGGAGCACAGAAGCAGAG
8          CTGGTGAGACTGAGGAATCAGGCTCATAAAGGGTTCTCAGGAGGGAGCACAGAAGCAGAG
9          CTGGTGAGACTGAGGAATCAGGCTCATAAAGGGTTCTCAGGAGGGAGCACAGAAGCAGAG
           ************************************************************

1          GAATTAAGGTCCAGTATGTCCATGGTGGGTCACCAGGCATGGCAGAACCTTACCACCTAG 1050
2          GAATTAAGGTCCAGTATGTCCATGGTGGGTCACCAGGCATGGCAGAACCTTACCACCTAG
3          GAATTAAGGTCCAGTATGTCCATGGTGGGTCACCAGGCATGGCAGAACCTTACCACCTAG
4          GAATTAAGGTCCAGTATGTCCATGGTGGGTCACCAGGCATGGCAGAACCTTACCACCTAG
5          GAATTAAGGTCCAGTATGTCCATGGTGGGTCACCAGGCATGGCAGAACCTTACCACCTAG
6          GAATTAAGGTCCAGTATGTCCATGGTGGGTCACCAGGCATGGCAGAACCTTACCACCTAG
7          GAATTAAGGTCCAGTATGTCCATGGTGGGTCACCAGGCATGGCAGAACCTTACCACCTAG
8          GAATTAAGGTCCAGTATGTCCATGGTGGGTCACCAGGCATGGCAGAACCTTACCACCTAG
9          GAATTAAGGTCCAGTATGTCCATGGTGGGTCACCAGGCATGGCAGAACCTTACCACCTAG
           ************************************************************

1          TCATTTACCCAGCAGGCCTGTCGGTATTCAGGAACACCTGCCCCCATCCCAGGAGAGAGC 1110
2          TCATTTACCCAGCAGGCCTGTCGGTATTCAGGAACACCTGCCCCCATCCCAGGAGAGAGC
3          TCATTTACCCAGCAGGCCTGTCGGTATTCAGGAACACCTGCCCCCATCCCAGGAGAGAGC
4          TCATTTACCCAGCAGGCCTGTCGGTATTCAGGAACACCTGCCCCCATCCCAGGAGAGAGC
5          TCATTTACCCAGCAGGCCTGTCGGTATTCAGGAACACCTGCCCCCATCCCAGGAGAGAGC
6          TCATTTACCCAGCAGGCCTGTCGGTATTCAGGAACACCTGCCCCCATCCCAGGAGAGAGC
7          TCATTTACCCAGCAGGCCTGTCGGTATTCAGGAACACCTGCCCCCATCCCAGGAGAGAGC
8          TCATTTACCCAGCAGGCCTGTCGGTATTCAGGAACACCTGCCCCCATCCCAGGAGAGAGC
9          TCATTTACCCAGCAGGCCTGTCGGTATTCAGGRACACCTGCCCCCATCCCAGGAGAGAGC
           ************************************************************

1          CCCATCCCTCCATC 1168
2          CCCATCCCTCCATC
3          CCCATCCCTCCATC
4          CCCATCCCTCCATC
5          CCCATCCCTCCATC
6          CCCATCCCTCCATC
7          CCCATCCCTCCATC
8          CCCATCCCTCCATC
9          CCCATCCCTCCATC
           **************
```

The database exonic sequence of 3B is shown in red (from position 630 to 1168) and is 539bp in size. The size of the PCR product is 539bp, which includes part of the exonic sequence from 3A, exon 3B and part of exon 3C. The overlapping exonic sequence of exon 3A and 3C is underlined.

1 = sequence of the exon from the database (http://www.ncbi.nlm.nih/nucleotide)

2 = forward sequence
3 = reverse complement sequence of the reverse } an unaffected individual

4 = forward sequence
5 = reverse complement sequence of the reverse } affected individual from pedigree 1

6 = forward sequence
7 = reverse complement sequence of the reverse } affected individual from pedigree 2

8 = forward sequence
9 = reverse complement sequence of the reverse } affected individual from pedigree 5

Table 6 CLUSTALW 1.81 alignment of exon 3C of *KIR2.4*

```
1                                                                 TCGCCTCCTGCTGCAGCCCCGTGTGA 1192
2                                                                 TCGCCTCCTGCTGCAGCCCCGTGTGA
3                                                                 TCGCCTCCTGCTGCAGCCCCGTGTGA
4                                                                 TCGCCTCCTGCTGCAGCCCCGTGTGA
5                                                                 TCCCCTCCTGGTGCAACCCCGGGTGA
6                                                                 TCGCCTCCTGCTGCAACCCCGTGTGA
7                                                                 TCGCCTCCTGCTGCAGCCCCGTGTGA
8                                                                 TCGCCTCCTGCTGCAGCCCCGTGTGA
9                                                                 TCGCCTCCTGCTGCAGCCCCGTGTGA
                                                                  ***************************

1                CCCCAGAGGGTGAGTACATCCCGCTGGACCACCAGGATGTGGATGTGGGCTTTGATGGAG 1252
2                CCCCAGAAGGTGAGTACATCCCGCTGGACCACCAGGATGTGGATGTGGGCTTTGATGGAG
3                CCCCAGAGGGTGAGTACATCCCGCTGGACCACCAGGATGTGGATGTGGGCTTTGATGGAG
4                CCCCAGAGGGTGAGTACATCCCGCTGGACCACCAGGATGTGGATGTGGGCTTTGATGGAG
5                CCCCAGAGGGTGAGTACATCCCGCTGGACCACCAAGATGTGAATGTGGGCTTTGATGGAG
6                CCCCAGAGGGTGAGTACATCCCGCTGGACCACCAGGATGTGGATGTGGGCTTTGATGGAG
7                CCCCAGAGGGTGAGTACATCCCGCTGGACCACCAGGATGTGGATGTGGGCTTTGATGGAG
8                CCCCAGAGGGTGAGTACATCCCGCTGGACCACCAGGATGTGGATGTGGGCTTTGATGGAG
9                CCCCAGAGGGTGAGTACATCCCGCTGGACCACCAGGATGTGGATGTGGGCTTTGATGGAG
                 ************************************************************

1                GCACCGATCGTATCTTCCTCGTGTCCCCCATCACCATCGTCCATGAGATCGACTCTGCCA 1312
2                GCACCGATCGTATCTTCCTCGTGTCCCCCATCACCATCGTCCATGAGATCGACTCTGCCA
3                GCACCGATCGTATCTTCCTCGTGTCCCCCATCACCATCGTCCATGAGATCGACTCTGCCA
4                GCACCGATCGTATCTTCCTCGTGTCCCCCATCACCATCGTCCATGAGATCGACTCTGCCA
5                GCACCGATCGTATCTTCCTCGTGTCCCCCATCACCATCGTCCATGAGATCGACTCTGCCA
6                GCACCGATCGTATCTTCCTCGTGTCCCCCATCACCATCGTCCATGAGATCGACTCTGCCA
7                GCACCGATCGTATCTTCCTCGTGTCCCCCATCACCATCGTCCATGAGATCGACTCTGCCA
8                GCACCGATCGTATCTTCCTCGTGTCCCCCATCACCATCGTCCATGAGATCGACTCTGCCA
9                GCACCGATCGTATCTTCCTCGTGTCCCCCATCACCATCGTCCATGAGATCGACTCTGCCA
                 ************************************************************

1                GTCCTCTGTATGAGCTAGGACGTGCCGAGCTGGCCAGGGCTGACTTTGAGCTGGTGGTC 1372
2                GTCCTCTGTATGAGCTAGGACGTGCCGAGCTGGCCAGGGCTGACTTTGAGCTGGTGGTC
3                GTCCTCTGTATGAGCTAGGACGTGCCGAGCTGGCCAGGGCTGACTTTGAGCTGGTGGTC
4                GTCCTCTGTATGAGCTAGGACGTGCCGAGCTGGCCAGGGCTGACTTTGAGCTGGTGGTC
5                GTCCTCTGTATGAGCTAGGACGTGCCGAGCTGGCCAGGGCTGACTTTGAGCTGGTGGTC
6                GTCCTCTGTATGAGCTAGGACGTGCCGAGCTGGCCAGGGCTGACTTTGAGCTGGTGGTC
7                GTCCTCTGTATGAGCTAGGACGTGCCGAGCTGGCCAGGGCTGACTTTGAGCTGGTGGTC
8                GTCCTCTGTATGAGCTAGGACGTGCCGAGCTGGCCAGGGCTGACTTTGAGCTGGTGGTC
9                GTCCTCTGTATGAGCTAGGACGTGCCGAGCTGGCCAGGGCTGACTTTGAGCTGGTGGTC
                 ************************************************************

1                ATTCTCGAGGGGATGGTTGAGGCCACAGCCATGACCACACAGTGTCGCTCGTCCTACCTC 1432
2                ATTCTCGAGGGGATGGTTGAGGCCACAGCCATGACCACACAGTGTCGCTCGTCCTACCTC
3                ATTCTCGAGGGGATGGTTGAGGCCACAGCCATGACCACACAGTGTCGCTCGTCCTACCTC
4                ATTCTCGAGGGGATGGTTGAGGCCACAGCCATGACCACACAGTGTCGCTCGTCCTACCTC
5                ATTCTCGAGGGGATGGTTGAGGCCACAGCCATGACCACACAGTGTCGCTCGTCCTACCTC
6                ATTCTCGAGGGGATGGTTGAGGCCACAGCCATGACCACACAGTGTCGCTCGTCCTACCTC
7                ATTCTCGAGGGGATGGTTGAGGCCACAGCCATGACCACACAGTGTCGCTCGTCCTACCTC
8                ATTCTCGAGGGGATGGTTGAGGCCACAGCCATGACCACACAGTGTCGCTCGTCCTACCTC
9                ATTCTCGAGGGGATGGTTGAGGCCACAGCCATGACCACACAGTGTCGCTCGTCCTACCTC
                 ************************************************************
```

```
1          CCTGGTGAACTGCTCTGGGGCCATCGTTTTGAGCCAGTTCTCTTCCAGCGTGGCTCCCAG 1492
2          CCTGGTGAACTGCTCTGGGGCCATCGTTTTGAGCCAGTTCTCTTCCAGCGTGGCTCCCAG
3          CCTGGTGAACTGCTCTGGGGCCATCGTTTTGAGCCAGTTCTCTTCCAGCGTGGCTCCCAG
4          CCTGGTGAACTGCTCTGGGGCCATCGTTTTGAGCCAGTTCTCTTCCAGCGTGGCTCCCAG
5          CCTGGTGAACTGCTCTGGGGCCATCGTTTTGAGCCAGTTCTCTTCCAGCGTGGCTCCCAG
6          CCTGGTGAACTGCTCTGGGGCCATCGTTTTGAGCCAGTTCTCTTCCAGCGTGGCTCCCAG
7          CCTGGTGAACTGCTCTGGGGCCATCGTTTTGAGCCAGTTCTCTTCCAGCGTGGCTCCCAG
8          CCTGGTGAACTGCTCTGGGGCCATCGTTTTGAGCCAGTTCTCTTCCAGCGTGGCTCCCAG
9          CCTGGTGAACTGCTCTGGGGCCATCGTTTTGAGCCAGTTCTCTTCCAGCGTGGCTCCCAG
           ************************************************************

1          TATGAGGTCGACTATCGCCACTTCCATCGCACTTATGAGGTCCCAGGGACACCGGTC     1549
2          TATGAGGTCGACTATCGCCACTTCCATCGCACTTATGAGGTCCCAGGGACACCGGTC
3          TATGAGGTCGACTATCGCCACTTCCATCGCACTTATGAGGTCCCAGGGACACCGGTC
4          TATGAGGTCGACTATCGCCACTTCCATCGCACTTATGAGGTCCCAGGGACACCGGTC
5          TATGAGGTCGACTATCGCCACTTCCATCGCACTTATGAGGTCCCAGGGACACCGGTC
6          TATGAGGTCGACTATCGCCACTTCCATCGCACTTATGAGGTCCCAGGGACACCGGTC
7          TATGAGGTCGACTATCGCCACTTCCATCGCACTTATGAGGTCCCAGGGACACCGGTC
8          TATGAGGTCGACTATCGCCACTTCCATCGCACTTATGAGGTCCCAGGGACACCGGTC
9          TATGAGGTCGACTATCGCCACTTCCATCGCACTTATGAGGTCCCAGGGACACCGGTC
           ************************************************************

1          TGCAGTGCTAAGGAGCTGGATGAACGGGCAGAGCAGGCTTCCCACAGCCTCAAGTCTAGT 1609
2          TGCAGTGCTAAGGAGCTGGATGAACGGGCAGAGCAGGCTTCCCACAGCCTCAAGTCTAGT
3          TGCAGTGCTAAGGAGCTGGATGAACGGGCAGAGCAGGCTTCCCACAGCCTCAAGTCTAGT
4          TGCAGTGCTAAGGAGCTGGATGAACGGGCAGAGCAGGCTTCCCACAGCCTCAAGTCTAGT
5          TGCAGTGTTAAGGAGCTGGATGAACGGGCAGAGCAGGCTTCCCACAGCCTCAAGTCTAGT
6          TGCAGTGCTAAGGAGCTGGATGAACGGGCAGAGCAGGCTTCCCACAGCCTCAAGTCTAGT
7          TGCAGTGCTAAGGAGCTGGATGAACGGGCAGAGCAGGCTTCCCACAGCCTCAAGTCTAGT
8          TGCAGTGCTAAGGAGCTGGATGAACGGGCAGAGCAGGCTTCCCACAGCCTCAAGTCTAGT
9          TGCAGTGCTAAGGAGCTGGATGAACGGGCAGAGCAGGCTTCCCACAGCCTCAAGTCTAGT
           ************************************************************

1          TTCCCCGGCTCTCTGACTGCATTTTGTTATGAGAATGAACTTGCTCTGAGCTGCTGCCAG 1669
2          TTCCCCGGCTCTCTGACTGCATTTTGTTATGAGAATGAACTTGCTCTGAGCTGCTGCCAG
3          TTCCCCGGCTCTCTGACTGCATTTTGTTATGAGAATGAACTTGCTCTGAGCTGCTGCCAG
4          TTCCCCGGCTCTCTGACTGCATTTTGTTATGAGAATGAACTTGCTCTGAGCTGCTGCCAG
5          TTCCCCGGCTCTCTGACTGCATTTTGTTATGAGAATGAACTTGCTCTGAGCTGCTGCCAG
6          TTCCCCGGCTCTCTGACTGCATTTTGTTATGAGAATGAACTTGCTCTGAGCTGCTGCCAG
7          TTCCCCGGCTCTCTGACTGCATTTTGTTATGAGAATGAACTTGCTCTGAGCTGCTGCCAG
8          TTCCCCGGCTCTCTGACTGCATTTTGTTATGAGAATGAACTTGCTCTGAGCTGCTGCCAG
9          TTCCCCGGCTCTCTGACTGCATTTTGTTATGAGAATGAACTTGCTCTGAGCTGCTGCCAG
           ************************************************************

1          GAGGAAGATGAGGACGATGAGACTGAGGAAGGGAATGGGGTGGAAACAGAAGATGGGGCT 1729
2          GAGGAAGATGAGGACGATGAGACTGAGGAAGGGAATGGGGTGGAAACAGAAGATGGGGCT
3          GAGGAAGATGAGGACGATGAGACTGAGGAAGGGAATGGGGTGGAAACAGAAGATGGGGCT
4          GAGGAAGATGAGGACGATGAGACTGAGGAAGGGAATGGGGTGGAAACAGAAGATGGGGCT
5          GAGGAAGATGAGGACGATGAGACTGAGGAAGGGAATGGGGTGGAAACAGAAGATGGGGCT
6          GAGGAAGATGAGGACGATGAGACTGAGGAAGGGAATGGGGTGGAAACAGAAGATGGGGCT
7          GAGGAAGATGAGGACGATGAGACTGAGGAAGGGAATGGGGTGGAAACAGAAGATGGGGCT
8          GAGGAAGATGAGGACGATGAGACTGAGGAAGGGAATGGGGTGGAAACAGAAGATGGGGCT
9          GAGGAAGATGAGGACGATGAGACTGAGGAAGGGAATGGGGTGGAAACAGAAGATGGGGCT
           ************************************************************
```

```
1          GCTAGCCCCCGAGTTCTCACACCAACCCTGGCGCTGACCCTGCCTCCATGATGCAAACT 1789
2          GCTAGCCCCCGAGTTCTCACACCAACCCTGGCGCTGACCCTGCCTCCATGATGCAAACT
3          GCTAGCCCCCGAGTTCTCACACCAACCCTGGCGCTGACCCTGCCTCCATGATGCAAACT
4          GCTAGCCCCCGAGTTCTCACACCAACCCTGGCGCTGACCCTGCCTCCATGATGCAAACT
5          GCTAGCCCCCGAGTTCTCACACCAACCCTGGCGCTGACCCTGCCTCCATGATGCAAACT
6          GCTAGCCCCCGAGTTCTCACACCAACCCTGGCGCTGACCCTGCCTCCATGATGCAAACT
7          GCTAGCCCCCGAGTTCTCACACCAACCCTGGCGCTGACCCTGCCTCCATGATGCAAACT
8          GCTAGCCCCCGAGTTCTCACACCAACCCTGGCGCTGACCCTGCCTCCATGATGCAAACT
9          GCTAGCCCCCGAGTTCTCACACCAACCCTGGCGCTGACCCTGCCTCCATGATGCAAACT
           ************************************************************
```

The database exonic sequence of 3C is shown in red (from position 1166 to 1741). The size of the PCR product is 624bp, which includes the overlapping exonic sequence of exon 3B with 3C, entire sequence of exon 3A and flanking the intronic sequence. The overlapping sequence is underlined.

1 = sequence of the exon from the database (http://www.ncbi.nlm.nih/nucleotide)

2 = forward sequence
3 = reverse complement sequence of the reverse } an unaffected individual

4 = forward sequence
5 = reverse complement sequence of the reverse } affected individual from pedigree 1

6 = forward sequence
7 = reverse complement sequence of the reverse } affected individual from pedigree 2

8 = forward sequence
9 = reverse complement sequence of the reverse } affected individual from pedigree 5

LIN-7B

Table 7 CLUSTALW 1.81 alignments of exon 1 of *LIN-7B*

```
1              CGCCAGGGCAGGCGGGCGGCTGGCAGCTGTGGCGCCGACATGGCTGCGCTGGTGGAGC 58
2              CGCCAGGGCAGGCGGGCGGCTGGCAGCTGTGGCGCCGACATGGCTGCGCTGGTGGAGC
3              CGCCAGGGCAGGCGGGCGGCTGGCAGCTGTGGCGCCGACATGGCTGCGCTGGTGGAGC
5              CGCCAGGGCAGGCGGGCGGCTGGCAGCTGTGGCGCCGACATGGCTGCGCTGGTGGAGC
6              CGCCAGGGCAGGCGGGCGGCTGGCAGCTGTGGCGCCGACATGGCTGCGCTGGTGGAGC
7              CGCCAGGGCAGGCGGGCGGCTGGCAGCTGTGGCGCCGACATGGCTGCGCTGGTGGAGC
               ************************************************************

1              CGCTGGGGCTGGAGCGGGGTAAGCGTGCGCC 89
2              CGCTGGGGCTGGAGCGGGGTAAGCGTGCGCC
3              CGCTGGGGCTGGAGCGGGGTAAGCGTGCGCC
5              CGCTGGGGCTGGAGCGGGGTAAGCGTGCGCC
6              CGCTGGGGCTGGAGCGGGGTAAGCGTGCGCC
7              CGCTGGGGCTGGAGCGGGGTAAGCGTGCGCC
               ********************************
```

The database exonic sequence is shown in red (from position 40 to 82) and is 43bp in size. The PCR product is 89bp in size, which includes the flanking intronic sequence and the entire exonic sequence.

1 = sequence of the exon from the database (http://www.ncbi.nlm.nih/nucleotide)

2 = forward sequence 3 = reverse complement sequence of the reverse	an unaffected individual
4 = forward sequence 5 = reverse complement sequence of the reverse	affected individual from pedigree 1
6 = forward sequence 7 = reverse complement sequence of the reverse	affected individual from pedigree 2
8 = forward sequence 9 = reverse complement sequence of the reverse	affected individual from pedigree 5

Table 8 CLUSTALW 1.81 alignments of exon 2 of *LIN-7B*

```
1                       GGTTTCTGCGCCCCGCCCCCGCCCCGCAGACGTGTCCCGGGCGGTTGAGCTC 52
2                       GGTTTCTGCGCCCCGCCCCCGCCCCGCAGACGTGTCCCGGGCGGTTGAGCTC
3                       GGTTTCTGCGCCCCGCCCCCGCCCCGCAGACGTGTCCCGGGCGGTTGAGCTC
4                       GGTTTCTGCGCCCCGCCCCCGCCCCGCAGACGTGTCCCGGGCGGTTGAGCTC
5                       GGTTTCTGCGCCCCGCCCCCGCCCCGCAGACGTGTCCCGGGCGGTTGAGCTC
8                       GGTTTCTGCGCCCCGCCCCCGCCCCGCAGACGTGTCCCGGGCGGTTGAGCTC
9                       GGTTTCTGCGCCCCGCCCCCGCCCCGCAGACGTGTCCCGGGCGGTTGAGCTC
                        ************************************************************

1               CTCGAGCGGCTCCAGCGCAGCGGGGAGCTGCCGCCGCAGAAGCTGCAGGCCCTCCAGCGA 112
2               CTCGAGCGGCTCCAGCGCAGCGGGGAGCTGCCGCCGCAGAAGCTGCAGGCCCTCCAGCGA
3               CTCGAGCGGCTCCAGCGCAGCGGGGAGCTGCCGCCGCAGAAGCTGCAGGCCCTCCAGCGA
4               CTCGAGCGGCTCCAGCGCAGCGGGGAGCTGCCGCCGCAGAAGCTGCAGGCCCTCCAGCGA
5               CTCGAGCGGCTCCAGCGCAGCGGGGAGCTGCCGCCGCAGAAGCTGCAGGCCCTCCAGCGA
8               CTCGAGCGGCTCCAGCGCAGCGGGGAGCTGCCGCCGCAGAAGCTGCAGGCCCTCCAGCGA
9               CTCGAGCGGCTCCAGCGCAGCGGGGAGCTGCCGCCGCAGAAGCTGCAGGCCCTCCAGCGA
                ************************************************************

1               GTTCTGCAGAGCCGCTTCTGCTCCGCTATCCGAGAGGTGAGGGGCGCGCGGCGCAGGGGC 172
2               GTTCTGCAGAGCCGCTTCTGCTCCGCTATCCGAGAGGTGAGGGGCGCGCGGCGCAGGGGC
3               GTTCTGCAGAGCCGCTTCTGCTCCGCTATCCGAGAGGTGAGGGGCGCGCGGCGCAGGGGC
4               GTTCTGCAGAGCCGCTTCTGCTCCGCTATCCGAGAGGTGAGGGGCGCGCGGCGCAGGGGC
5               GTTCTGCAGAGCCGCTTCTGCTCCGCTATCCGAGAGGTGAGGGGCGCGCGGCGCAGGGGC
8               GTTCTGCAGAGCCGCTTCTGCTCCGCTATCCGAGAGGTGAGGGGCGCGCGGCGCAGGGGC
9               GTTCTGCAGAGCCGCTTCTGCTCCGCTATCCGAGAGGTGAGGGGCGCGCGGCGCAGGGGC
                ************************************************************
```

The database exonic sequence is shown in red (from position 30 to 148) and is 119bp in size. The size of the PCR product is 172bp, which includes the flanking intronic sequence and the entire exonic sequence.

1 = sequence of the exon from the database (http://www.ncbi.nlm.nih/nucleotide)

2 = forward sequence 3 = reverse complement sequence of the reverse	an unaffected individual
4 = forward sequence 5 = reverse complement sequence of the reverse	affected individual from pedigree 1
6 = forward sequence 7 = reverse complement sequence of the reverse	affected individual from pedigree 2
8 = forward sequence 9 = reverse complement sequence of the reverse	affected individual from pedigree 5

Table 9 CLUSTALW 1.81 alignments of exon 3 of *LIN-7B*

```
1                                                                CGGGGGCGGGGCGGATCCTTGCGTG 25
2                                                                CGGGGGCGGGGCGGATCCTTGCGTG
3                                                                CGGGGGCGGGGCGGATCCTTGCGTG
4                                                                CGGGGGCGGGGCGGATCCTTGCGTG
5                                                                CGGGGGCGGGGCGGATCCTTGCGTG
6                                                                CGGGGGCGGGGCGGATCCTTGCGTG
8                                                                CGGGGGCGGGGCGGATCCTTGCGTG
9                                                                CGGGGGCGGGGCGGATCCTTGCGTG
                                                                 *************************

1                GTAGCGGGAGAGGGTCTAGAGGCCTGAACTCCTGCGTCTAGGGCTGGAGGAGCTGGCGA 84
2                GTAGCGGGAGAGGGTCTAGAGGCCTGAACTCCTGCGTCTAGGGCTGGAGGAGCTGGCGA
3                GTAGCGGGAGAGGGTCTAGAGGCCTGAACTCCTGCGTCTAGGGCTGGAGGAGCTGGCGA
4                GTAGCGGGAGAGGGTCTAGAGGCCTGAACTCCTGCGTCTAGGGCTGGAGGAGCTGGCGA
5                GTAGCGGGAGAGGGTCTAGAGGCCTGAACTCCTGCGTCTAGGGCTGGAGGAGCTGGCGA
6                GTAGCGGGAGAGGGTCTAGAGGCCTGAACTCCTGCGTCTAGGGCTGGAGGAGCTGGCGA
8                GTAGCGGGAGAGGGTCTAGAGGCCTGAACTCCTGCGTCTAGGGCTGGAGGAGCTGGCGA
9                GTAGCGGGAGAGGGTCTAGAGGCCTGAACTCCTGCGTCTAGGGCTGGAGGAGCTGGCGA
                 ************************************************************

1                CTCCCTGATGCCGCTGCCTCCTCACCCAGGTGTATGAGCAGCTTTATGACACGCTGGACA 144
2                CTCCCTGATGCCGCTGCCTCCTCACCCAGGTGTATGAGCAGCTTTATGACACGCTGGACA
3                CTCCCTGATGCCGCTGCCTCCTCACCCAGGTGTATGAGCAGCTTTATGACACGCTGGACA
4                CTCCCTGATGCCGCTGCCTCCTCACCCAGGTGTATGAGCAGCTTTATGACACGCTGGACA
5                CTCCCTGATGCCGCTGCCTCCTCACCCAGGTGTATGAGCAGCTTTATGACACGCTGGACA
6                CTCCCTGATGCCGCTGCCTCCTCACCCAGGTGTATGAGCAGCTTTATGACACGCTGGACA
8                CTCCCTGATGCCGCTGCCTCCTCACCCAGGTGTATGAGCAGCTTTATGACACGCTGGACA
9                CTCCCTGATGCCGCTGCCTCCTCACCCAGGTGTATGAGCAGCTTTATGACACGCTGGACA
                 ************************************************************

1                TCACCGGCAGCGCCGAGATCCGAGCCCATGCCACAGCCAAGGTGGGCCCCGCACCCCATT 204
2                TCACCGGCAGCGCCGAGATCCGAGCCCATGCCACAGCCAAGGTGGGCCCCGCACCCCATT
3                TCACCGGCAGCGCCGAGATCCGAGCCCATGCCACAGCCAAGGTGGGCCCCGCACCCCATT
4                TCACCGGCAGCGCCGAGATCCGAGCCCATGCCACAGCCAAGGTGGGCCCCGCACCCCATT
5                TCACCGGCAGCGCCGAGATCCGAGCCCATGCCACAGCCAAGGTGGGCCCCGCACCCCATT
6                TCACCGGCAGCGCCGAGATCCGAGCCCATGCCACAGCCAAGGTGGGCCCCGCACCCCATT
8                TCACCGGCAGCGCCGAGATCCGAGCCCATGCCACAGCCAAGGTGGGCCCCGCACCCCATT
9                TCACCGGCAGCGCCGAGATCCGAGCCCATGCCACAGCCAAGGTGGGCCCCGCACCCCATT
                 ************************************************************

1                GCTCCTGGTGTCTATTTAACTGCCTAGTCGAACTCAATAT 244
2                GCTCCTGGTGTCTATTTAACTGCCTAGTCGAACTCAATAT
3                GCTCCTGGTGTCTATTTAACTGCCTAGTCGAACTCAATAT
4                GCTCCTGGTGTCTATTTAACTGCCTAGTCGAACTCAATAT
5                GCTCCTGGTGTCTATTTAACTGCCTAGTCGAACTCAATAT
6                GCTCCTGGTGTCTATTTAACTGCCTAGTCGAACTCAATAT
8                GCTCCTGGTGTCTATTTAACTGCCTAGTCGAACTCAATAT
9                GCTCCTGGTGTCTATTTAACTGCCTAGTCGAACTCAATAT
                 ****************************************
```

The database exonic sequence is shown in red (from position 30 to 230) and is 201bp in size. The size of the PCR product is 244bp, which includes the flanking intronic sequence and the entire exonic sequence.

1 = sequence of the exon from the database (http://www.ncbi.nlm.nih/nucleotide)

2 = forward sequence
3 = reverse complement sequence of the reverse } an unaffected individual

4 = forward sequence
5 = reverse complement sequence of the reverse } affected individual from pedigree 1

6 = forward sequence
7 = reverse complement sequence of the reverse } affected individual from pedigree 2

8 = forward sequence
9 = reverse complement sequence of the reverse } affected individual from pedigree 5

Table 10 CLUSTALW 1.81 alignments of exon 4 of *LIN-7B*

```
1                    ATCCCCCACCCCAGCTTTCATGCCTACCTGGAAGGGGCCCTCATCTTCAGTCGC 54
2                    ATCCCCCACCCCAGCTTTCATGCCTACCTGGAAGGGGCCCTCATCTTCAGTCGC
3                    ATCCCCCACCCCAGCTTTCATGCCTACCTGGAAGGGGCCCTCATCTTCAGTCGC
4                    ATCCCCCACCCCAGCTTTCATGCCTACCTGGAAGGGGCCCTCATCTTCAGTCGC
5                    ATCCCCCACCCCAGCTTTCATGCCTACCTGGAAGGGGCCCTCATCTTCAGTCGC
6                    ATCCCCCACCCCAGCTTTCATGCCTACCTGGAAGGGGCCCTCATCTTCAGTCGC
7                    ATCCCCCACCCCAGCTTTCATGCCTACCTGGAAGGGGCCCTCATCTTCAGTCGC
8                    ATCCCCCACCCCAGCTTTCATGCCTACCTGGAAGGGGCCCTCATCTTCAGTCGC
9                    ATCCCCCACCCCAGCTTTCATGCCTACCTGGAAGGGGCCCTCATCTTCAGTCGC
                     ********************************************************

1                TTTCTCTCTCCCAGGCCACAGTGGCTGCCTTCACAGCCAGCGAGGGCCACGCACATCCCA 114
2                TTTCTCTCTCCCAGGCCACAGTGGCTGCCTTCACAGCCAGCGAGGGCCACGCACATCCCA
3                TTTCTCTCTCCCAGGCCACAGTGGCTGCCTTCACAGCCAGCGAGGGCCACGCACATCCCA
4                CTTCTCTCTCCCAGGCCACAGTGGCTGCCTTCACAGCCAGCGAGGGCCACGCACATCCCA
5                TTTCTCTCTCCCAGGCCACAGTGGCTGCCTTCACAGCCAGCGAGGGCCACGCACATCCCA
6                TTTCTCTCTCCCAGGCCACAGTGGCTGCCTTCACAGCCAGCGAGGGCCACGCACATCCCA
7                TTTCTCTCTCCCAGGCCACAGTGGCTGCCTTCACAGCCAGCGAGGGCCACGCACATCCCA
8                TTTCTCTCTCCCAGGCCACAGTGGCTGCCTTCACAGCCAGCGAGGGCCACGCACATCCCA
9                TTTCTCTCTCCCAGGCCACAGTGGCTGCCTTCACAGCCAGCGAGGGCCACGCACATCCCA
                 ************************************************************

1                GGGTAGTGGAGCTACCCAAGACGGATGAGGGCCTAGGCTTCAACATCATGGGTGGCAAAG 174
2                GGGTAGTGGAGCTACCCAAGACGGATGAGGGCCTAGGCTTCAACATCATGGGTGGCAAAG
3                GGGTAGTGGAGCTACCCAAGACGGATGAGGGCCTAGGCTTCAACATCATGGGTGGCAAAG
4                GGGTAGTGGAGCTACCCAAGACGGATGAGGGCCTAGGCTTCAACATCATGGGTGGCAAAG
5                GGGTAGTGGAGCTACCCAAGACGGATGAGGGCCTAGGCTTCAACATCATGGGTGGCAAAG
6                GGGTAGTGGAGCTACCCAAGACGGATGAGGGCCTAGGCTTCAACATCATGGGTGGCAAAG
7                GGGTAGTGGAGCTACCCAAGACGGATGAGGGCCTAGGCTTCAACATCATGGGTGGCAAAG
8                GGGTAGTGGAGCTACCCAAGACGGATGAGGGCCTAGGCTTCAACATCATGGGTGGCAAAG
9                GGGTAGTGGAGCTACCCAAGACGGATGAGGGCCTAGGCTTCAACATCATGGGTGGCAAAG
                 ************************************************************

1                AGCAAAACTCGCCCATCTACATCTCCCGGGTCATCCCAGGGGGGTGTGGCTGACCGCCATG 234
2                AGCAAAACTCGCCCATCTACATCTCCCGGGTCATCCCAGGGGGGTGTGGCTGACCGCCATG
3                AGCAAAACTCGCCCATCTACATCTCCCGGGTCATCCCAGGGGGGTGTGGCTGACCGCCATG
4                AGCAAAACTCGCCCATCTACATCTCCCGGGTCATCCCAGGGGGGTGTGGCTGACCGCCATG
5                AGCAAAACTCGCCCATCTACATCTCCCGGGTCATCCCAGGGGGGTGTGGCTGACCGCCATG
6                AGCAAAACTCGCCCATCTACATCTCCCGGGTCATCCCAGGGGGGTGTGGCTGACCGCCATG
7                AGCAAAACTCGCCCATCTACATCTCCCGGGTCATCCCAGGGGGGTGTGGCTGACCGCCATG
8                AGCAAAACTCGCCCATCTACATCTCCCGGGTCATCCCAGGGGGGTGTGGCTGACCGCCATG
9                AGCAAAACTCGCCCATCTACATCTCCCGGGTCATCCCAGGGGGGTGTGGCTGACCGCCATG
                 ************************************************************

1                GAGGCCTCAAGCGTGGGGATCAACTGTTGTCGGTGAACGGTGTGGTGAGTGGAGGGCT 292
2                GAGGCCTCAAGCGTGGGGATCAACTGTTGTCGGTGAACGGTGTGGTGAGTGGAGGGCT
3                GAGGCCTCAAGCGTGGGGATCAACTGTTGTCGGTGAACGGTGTGGTGAGTGGAGGGCT
4                GAGGCCTCAAGCGTGGGGATCAACTGTTGTCGGTGAACGGTGTGGTGAGTGGAGGGCT
5                GAGGCCTCAAGCGTGGGGATCAACTGTTGTCGGTGAACGGTGTGGTGAGTGGAGGGCT
6                GAGGCCTCAAGCGTGGGGATCAACTGTTGTCGGTGAACGGTGTGGTGAGTGGAGGGCT
7                GAGGCCTCAAGCGTGGGGATCAACTGTTGTCGGTGAACGGTGTGGTGAGTGGAGGGCT
8                GAGGCCTCAAGCGTGGGGATCAACTGTTGTCGGTGAACGGTGTGGTGAGTGGAGGGCT
9                GAGGCCTCAAGCGTGGGGATCAACTGTTGTCGGTGAACGGTGTGGTGAGTGGAGGGCT
                 ************************************************************
```

```
1                GAGGCAGGACTGGGGGA 309
2                GAGGCAGGACTGGGGGA
3                GAGGCAGGACTGGGGGA
4                GAGGCAGGACTGGGGGA
5                GAGGCAGGACTGGGGGA
6                GAGGCAGGACTGGGGGA
7                GAGGCAGGACTGGGGGA
8                GAGGCAGGACTGGGGGA
9                GAGGCAGGACTGGGGGA
                 *****************
```

The database exonic sequence is shown in red (from position 15 to 293) and is 279bp in size. The size of the PCR product is 309bp, which includes the flanking intronic sequence and the entire exonic sequence.

1 = sequence of the exon from the database (http://www.ncbi.nlm.nih/nucleotide)

2 = forward sequence
3 = reverse complement sequence of the reverse } an unaffected individual

4 = forward sequence
5 = reverse complement sequence of the reverse } affected individual from pedigree 1

6 = forward sequence
7 = reverse complement sequence of the reverse } affected individual from pedigree 2

8 = forward sequence
9 = reverse complement sequence of the reverse } affected individual from pedigree 5

.
.

Table 11 CLUSTALW 1.81 alignments of exon 5 of *LIN-7B*

```
1                                                                   AGTGGG  6
2                                                                   AGTGGG
3                                                                   AGTGGG
4                                                                   AGTGGG
5                                                                   AGTGGG
6                                                                   AGTGGG
7                                                                   AGTGGG
8                                                                   AGTGGG
9                                                                   AGTGGG
                                                                    ******

1           GGCTGGACGAGGGCAGCGGGCCCCAGGCTCAGCTGTCTGTGTTGGGCCCTGCAGAGCGTT  66
2           GGCTGGAGGAGGGCAGCGGGCCCCAGGCTCAGCTGTCTGTGTTGGGCCCTGCAGAGCGTT
3           GGCTGGACGAGGGCAGCGGGCCCCAGGCTCAGCTGTCTGTGTTGGGCCCTGCAGAGCGTT
4           GGCTGGACGAGGGCAGCGGGCCCCAGGCTCAGCTGTCTGTGTTGGGCCCTGCAGAGCGTT
5           GGCTGGACGAGGGCAGCGGGCCCCAGGCTCAGCTGTCTGTGTTGGGCCCTGCAGAGCGTT
6           GGCTGGACGAGGGCAGCGGGCCCCAGGCTCAGCTGTCTGTGTTGGGCCCTGCAGAGCGTT
7           GGCTGGACGAGGGCAGCGGGCCCCAGGCTCAGCTGTCTGTGTTGGGCCCTGCAGAGCGTT
8           GGCTGGACGAGGGCAGCGGGCCCCAGGCTCAGCTGTCTGTGTTGGGCCCTGCAGAGCGTT
9           GGCTGGACGAGGGCAGCGGGCCCCAGGCTCAGCTGTCTGTGTTGGGCCCTGCAGAGCGTT
            ************************************************************

1           GAGGGTGAGCAGCATGAGAAGGCGGTGGAGCTGCTGAAGGCGGCCCAGGGCTCGGTGAAG 126
2           GAGGGTGAGCAGCATGAGAAGGCGGTGGAGCTGCTGAAGGCGGCCCAGGGCTCGGTGAAG
3           GAGGGTGAGCAGCATGAGAAGGCGGTGGAGCTGCTGAAGGCGGCCCAGGGCTCGGTGAAG
4           GAGGGTGAGCAGCATGAGAAGGCGGTGGAGCTGCTGAAGGCGGCCCAGGGCTCGGTGAAG
5           GAGGGTGAGCAGCATGAGAAGGCGGTGGAGCTGCTGAAGGCGGCCCAGGGCTCGGTGAAG
6           GAGGGTGAGCAGCATGAGAAGGCGGTGGAGCTGCTGAAGGCGGCCCAGGGCTCGGTGAAG
7           GAGGGTGAGCAGCATGAGAAGGCGGTGGAGCTGCTGAAGGCGGCCCAGGGCTCGGTGAAG
8           GAGGGTGAGCAGCATGAGAAGGCGGTGGAGCTGCTGAAGGCGGCCCAGGGCTCGGTGAAG
9           GAGGGTGAGCAGCATGAGAAGGCGGTGGAGCTGCTGAAGGCGGCCCAGGGCTCGGTGAAG
            ************************************************************

1           CTGGTTGTCCGTTACACACCGCGAGTGCTGGAGGAGATGGAGGCCCGGTTCGAGAAGATG 186
2           CTGGTTGTCCGTTACACACCGCGAGTGCTGGAGGAGATGGAGGCCCGGTTCGAGAAGATG
3           CTGGTTGTCCGTTACACACCGCGAGTGCTGGAGGAGATGGAGGCCCGGTTCGAGAAGATG
4           CTGGTTGTCCGTTACACACCGCGAGTGCTGGAGGAGATGGAGGCCCGGTTCGAGAAGATG
5           CTGGTTGTCCGTTACACACCGCGAGTGCTGGAGGAGATGGAGGCCCGGTTCGAGAAGATG
6           CTGGTTGTCCGTTACACACCGCGAGTGCTGGAGGAGATGGAGGCCCGGTTCGAGAAGATG
7           CTGGTTGTCCGTTACACACCGCGAGTGCTGGAGGAGATGGAGGCCCGGTTCGAGAAGATG
8           CTGGTTGTCCGTTACACACCGCGAGTGCTGGAGGAGATGGAGGCCCGGTTCGAGAAGATG
9           CTGGTTGTCCGTTACACACCGCGAGTGCTGGAGGAGATGGAGGCCCGGTTCGAGAAGATG
            ************************************************************

1           CGCTCTGCCCGCCGGCGCCAACAGCATCAGAGCTACTCGTGAGCCCCTGGGTCACCACA 246
2           CGCTCTGCCCGCCGGCGCCAACAGCATCAGAGCTACTCGTGAGCCCCTGGGTCACCACA
3           CGCTCTGCCCGCCGGCGCCAACAGCATCAGAGCTACTCGTGAGCCCCTGGGTCACCACA
4           CGCTCTGCCCGCCGGCGCCAACAGCATCAGAGCTACTCGTGAGCCCCTGGGTCACCACA
5           CGCTCTGCCCGCCGGCGCCAACAGCATCAGAGCTACTCGTGAGCCCCTGGGTCACCACA
6           CGCTCTGCCCGCCGGCGCCAACAGCATCAGAGCTACTCGTGAGCCCCTGGGTCACCACA
7           CGCTCTGCCCGCCGGCGCCAACAGCATCAGAGCTACTCGTGAGCCCCTGGGTCACCACA
8           CGCTCTGCCCGCCGGCGCCAACAGCATCAGAGCTACTCGTGAGCCCCTGGGTCACCACA
9           CGCTCTGCCCGCCGGCGCCAACAGCATCAGAGCTACTCGTGAGCCCCTGGGTCACCACA
            ************************************************************
```

```
1          CCCCTGGGGCCTCCACGGGCTCCCTTTAACCCCAGG 282
2          CCCCTGGGGCCTCCACGGGCTCCCTTTAACCCCAGG
3          CCCCTGGGGCCTCCACGGGCTCCCTTTAACCCCAGG
4          CCCCTGGGGCCTCCACGGGCTCCCTTTAACCCCAGG
5          CCCCTGGGGCCTCCACGGGCTCCCTTTAACCCCAGG
6          CCCCTGGGGCCTCCACGGGCTCCCTTTAACCCCAGG
7          CCCCTGGGGCCTCCACGGGCTCCCTTTAACCCCAGG
8          CCCCTGGGGCCTCCACGGGCTCCCTTTAACCCCAGG
9          CCCCTGGGGCCTCCACGGGCTCCCTTTAACCCCAGG
           ************************************
```

The database exonic sequence is shown in red (from position 61 to 224) and is 164bp in size. The size of the PCR product is 282bp, which includes the flanking intronic sequence and the entire exonic sequence.

1 = sequence of the exon from the database (http://www.ncbi.nlm.nih/nucleotide)

2 = forward sequence
3 = reverse complement sequence of the reverse } an unaffected individual
4 = forward sequence
5 = reverse complement sequence of the reverse } affected individual from pedigree 1
6 = forward sequence
7 = reverse complement sequence of the reverse } affected individual from pedigree 2
8 = forward sequence
9 = reverse complement sequence of the reverse } affected individual from pedigree 5

.
.

BAX

Table 12 CLUSTALW 1.81 analysis of exon 1 of *BAX*

```
1                                                    GGGCGCGCTGCGGCTCGCCCGCGCGG 26
2                                                    GGGCGCGCTGCGGCTCGCCCGCGCGG
3                                                    GGGCGCGCTGCGGCTCGCCCGCGCGG
4                                                    GGGCGCGCTGCGGCTCGCCCGCGCGG
5                                                    GGGCGCGCTGCGGCTCGCCCGCGCGG
6                                                    GGGCGCGCTGCGGCTCGCCCGCGCGG
7                                                    GGGCGCGCTGCGGCTCGCCCGCGCGG
8                                                    GGGCGCGCTGCGGCTCGCCCGCGCGG
9                                                    GGGCGCGCTGCGGCTCGCCCGCGCGG
                                                     **************************

1              ACCCGGCGAGAGGCGGCGGCGGGAGCGGCGGTGTATGGACGGGTCCGGGGAGCAGCCCA 85
2              ACCCGGCGAGAGGCGGCGGCGGGAGCGGCGGTGTATGGACGGGTCCGGGGAGCAGCCCA
3              ACCCGGCGAGAGGCGGCGGCGGGAGCGGCGGTGTATGGACGGGTCCGGGGAGCAGCCCA
4              ACCCGGCGAGAGGCGGCGGCGGGAGCGGCGGTGTATGGACGGGTCCGGGGAGCAGCCCA
5              ACCCGGCGAGAGGCGGCGGCGGGAGCGGCGGTGTATGGACGGGTCCGGGGAGCAGCCCA
6              ACCCGGCGAGAGGCGGCGGCGGGAGCGGCGGTGTATGGACGGGTCCGGGGAGCAGCCCA
7              ACCCGGCGAGAGGCGGCGGCGGGAGCGGCGGTGTATGGACGGGTCCGGGGAGCAGCCCA
8              ACCCGGCGAGAGGCGGCGGCGGGAGCGGCGGTGTATGGACGGGTCCGGGGAGCAGCCCA
9              ACCCGGCGAGAGGCGGCGGCGGGAGCGGCGGTGTATGGACGGGTCCGGGGAGCAGCCCA
               ************************************************************

1              GAGGCGGGGGTGAGGCGGGAGGCAGACGGG 115
2              GAAGCGGGGGTGAGGCGGGAGGCAGACGGG
3              GAGGCGGGGGTGAGGCGGGAGGCAGACGGG
4              GAGGCGGGGGTGAGGCGGGAGGCCGACGGG
5              GAGGCGGGGGTGAGGCGGGAGGCAGACGGG
6              GANGCGAGGGTGAGGCGGGAGGCAGACGGG
7              GAGGCGGGGGTGAGGCGGGAGGCAGACGGG
8              GAGGCGGGGGTGAGGCGGGAGGCAGACGGG
9              GAGGCGGGGGTGAGGCGGGAGGCAGACGGG
               ********************************
```

The database exonic sequence is shown in red (from position 61 to 94) and is 34bp in size. The size of the PCR product is 115bp, which includes the flanking intronic sequence and the entire exonic sequence.

1 = sequence of the exon from the database (http://www.ncbi.nlm.nih/nucleotide)

2 = forward sequence
3 = reverse complement sequence of the reverse } an unaffected individual
4 = forward sequence
5 = reverse complement sequence of the reverse } affected individual from pedigree 1
6 = forward sequence
7 = reverse complement sequence of the reverse } affected individual from pedigree 2
8 = forward sequence
9 = reverse complement sequence of the reverse } affected individual from pedigree 5

Table 13 CLUSTALW 1.81 analysis of exon 2 and 3 of *BAX*.

```
1                                                                 TTCCCTTCCTTTCTCCTCTAGGGCCCAC 28
2                                                                 TTCCCTTCCTTTCTCCTCTAGGGCCCAC
3                                                                 TTCCCTTCCTTTCTCCTCTAGGGCCCAC
4                                                                 TTCCCTTCCTTTCTCCTCTAGGGCCCAC
6                                                                 TTCCCTTCCTTTCTCCTCTAGGGCCCAC
7                                                                 TTCCCTTCCTTTCTCCTCTAGGGCCCAC
8                                                                 TTCCCTTCCTTTCTCCTCTAGGGCCCAC
9                                                                 TTCCCTTCCTTTCTCCTCTAGGGCCCAC
                                                                  ****************************

1                CAGCTCTGAGCAGATCATGAAGACAGGGGCCCTTTTGCTTCAGGGGTGAGTTTGAGG 85
2                CAGCTCTGAGCAGATCATGAAGACAGGGGCCCTTTTGCTTCAGGGGTGAGTTTGAGG
3                CAGCTCTGAGCAGATCATGAAGACAGGGGCCCTTTTGCTTCAGGGGTGAGTTTGAGG
4                CAGCTCTGAGCAGATCATGAAGACAGGGGCCCTTTTGCTTCAGGGGTGAGTTTGAGG
6                CAGCTCTGAGCAGATCATGAAGACAGGGGCCCTTTTGCTTCAGGGGTGAGTTTGAGG
7                CAGCTCTGAGCAGATCATGAAGACAGGGGCCCTTTTGCTTCAGGGGTGAGTTTGAGG
8                CAGCTCTGAGCAGATCATGAAGACAGGGGCCCTTTTGCTTCAGGGGTGAGTTTGAGG
9                CAGCTCTGAGCAGATCATGAAGACAGGGGCCCTTTTGCTTCAGGGGTGAGTTTGAGG
                 ************************************************************

1                TCTGATTATTGTGGCACAGATTTGAGAGAGTGACACCCCGTTCTGATTCTGCACCCTCAC 145
2                TCTGATTATTGTGGCACAGATTTGAGAGAGTGACACCCCGTTCTGATTCTGCACCCTCAC
3                TCTGATTATTGTGGCACAGATTTGAGAGAGTGACACCCCGTTCTGATTCTGCACCCTCAC
4                TCTGATTATTGTGGCACAGATTTGAGAGAGTGACACCCCGTTCTGATTCTGCACCCTCAC
6                TCTGATTATTGTGGCACAGATTTGAGAGAGTGACACCCCGTTCTGATTCTGCACCCTCAC
7                TCTGATTATTGTGGCACAGATTTGAGAGAGTGACACCCCGTTCTGATTCTGCACCCTCAC
8                TCTGATTATTGTGGCACAGATTTGAGAGAGTGACACCCCGTTCTGATTCTGCACCCTCAC
9                TCTGATTATTGTGGCACAGATTTGAGAGAGTGACACCCCGTTCTGATTCTGCACCCTCAC
                 ************************************************************

1                TCCATCCCCACTCTAGTTTCATCCAGGATCGAGCAGGGCGAATGGGGGGGGGAGGCACCCG 205
2                TCCATCCCCACTCTAGTTTCATCCAGGATCGAGCAGGGCGAATGGGGGGGGGAGGCACCCG
3                TCCATCCCCACTCTAGTTTCATCCAGGATCGAGCAGGGCGAATGGGGGGGGGAGGCACCCG
4                TCCATCCCCACTCTAGTTTCATCCAGGATCGAGCAGGGCGAATGGGGGGGGGAGGCACCCG
6                TCCATCCCCACTCTAGTTTCATCCAGGATCGAGCAGGGCGAATGGGGGGGGGAGGCACCCG
7                TCCATCCCCACTCTAGTTTCATCCAGGATCGAGCAGGGCGAATGGGGGGGGGAGGCACCCG
8                TCCATCCCCACTCTAGTTTCATCCAGGATCGAGCAGGGCGAATGGGGGGGGGAGGCACCCG
9                TCCATCCCCACTCTAGTTTCATCCAGGATCGAGCAGGGCGAATGGGGGGGGGAGGCACCCG
                 ************************************************************

1                AGCTGGCCCTGGACCCGGTGCCTCAGGATGCGTCCACCAAGAAGCTGAGCGAGTGTCT 263
2                AGCTGGCCCTGGACCCGGTGCCTCAGGATGCGTCCACCAAGAAGCTGAGCGAGTGTCT
3                AGCTGGCCCTGGACCCGGTGCCTCAGGATGCGTCCACCAAGAAGCTGAGCGAGTGTCT
4                AGCTGGCCCTGGACCCGGTGCCTCAGGATGCGTCCACCAAGAAGCTGAGCGAGTGTCT
6                AGCTGGCCCTGGACCCGGTGCCTCAGGATGCGTCCACCAAGAAGCTGAGCGAGTGTCT
7                AGCTGGCCCTGGACCCGGTGCCTCAGGATGCGTCCACCAAGAAGCTGAGCGAGTGTCT
8                AGCTGGCCCTGGACCCGGTGCCTCAGGATGCGTCCACCAAGAAGCTGAGCGAGTGTCT
9                AGCTGGCCCTGGACCCGGTGCCTCAGGATGCGTCCACCAAGAAGCTGAGCGAGTGTCT
                 ************************************************************

1                CAAGCGCATCGGGGACGAACTGGACAGTAACATGGAGCTGCAGAGGTGTGGGCCCCTGGG 323
2                CAAGCGCATCGGGGACGAACTGGACAGTAACATGGAGCTGCAGAGGTGTGGGCCCCTGGG
3                CAAGCGCATCGGGGACGAACTGGACAGTAACATGGAGCTGCAGAGGTGTGGGCCCCTGGG
4                CAAGCGCATCGGGGACGAACTGGACAGTAACATGGAGCTGCAGAGGTGTGGGCCCCTGGG
6                CAAGCGCATCGGGGACGAACTGGACAGTAACATGGAGCTGCAGAGGTGTGGGCCCCTGGG
7                CAAGCGCATCGGGGACGAACTGGACAGTAACATGGAGCTGCAGAGGTGTGGGCCCCTGGG
8                CAAGCGCATCGGGGACGAACTGGACAGTAACATGGAGCTGCAGAGGTGTGGGCCCCTGGG
9                CAAGCGCATCGGGGACGAACTGGACAGTAACATGGAGCTGCAGAGGTGTGGGCCCCTGGG
                 ************************************************************
```

The exonic sequence of exon 2 (position 22-73, which is 52bp in size) and exon 3 (position 162-310, which is 149bp in size) is shown in red. The size of the PCR product containing both exons 2 and 3 is 323bp and is flanked by the intronic sequences.

1 = sequence of the exon from the database (http://www.ncbi.nlm.nih/nucleotide)

2 = forward sequence 3 = reverse complement sequence of the reverse	an unaffected individual
4 = forward sequence 5 = reverse complement sequence of the reverse	affected individual from pedigree 1
6 = forward sequence 7 = reverse complement sequence of the reverse	affected individual from pedigree 2
8 = forward sequence 9 = reverse complement sequence of the reverse	affected individual from pedigree 5

Table 14 CLUSTALW 1.81 analysis of exon 4 of *BAX*.

```
1                                                                   ACTCC 5
2                                                                   ACTCC
3                                                                   ACTCC
4                                                                   ACTCC
5                                                                   ACTCC
6                                                                   ACTCC
7                                                                   ACTCC
8                                                                   ACTCC
9                                                                   ACTCC
                                                                    *****

1                    CCCCGAGAGGTCTTTTTCCGAGTGGCAGCTGACATGTTTTCTGACGGCAACT 57
2                    CCCCGAGAGGTCTTTTTCCGAGTGGCAGCTGACATGTTTTCTGACGGCAACT
3                    CCCCGAGAGGTCTTTTTCCGAGTGGCAGCTGACATGTTTTCTGACGGCAACT
4                    CCCCGAGAGGTCTTTTTCCGAGTGGCAGCTGACATGTTTTCTGACGGCAACT
5                    CCCCGAGAGGTCTTTTTCCGAGTGGCAGCTGACATGTTTTCTGACGGCAACT
6                    CCCCGAGAGGTCTTTTTCCGAGTGGCAGCTGACATGTTTTCTGACGGCAACT
7                    CCCCGAGAGGTCTTTTTCCGAGTGGCAGCTGACATGTTTTCTGACGGCAACT
8                    CCCCGAGAGGTCTTTTTCCGAGTGGCAGCTGACATGTTTTCTGACGGCAACT
9                    CCCCGAGAGGTCTTTTTCCGAGTGGCAGCTGACATGTTTTCTGACGGCAACT
                     ************************************************************

1              TCAACTGGGGCCGGGTTGTCGCCCTTTTCTACTTTGCCAGCAAACTGGTGCTCAAGGT 115
2              TCAACTGGGGCCGGGTTGTCGCCCTTTTCTACTTTGCCAGCAAACTGGTGCTCAAGGT
3              TCAACTGGGGCCGGGTTGTCGCCCTTTTCTACTTTGCCAGCAAACTGGTGCTCAAGGT
4              TCAACTGGGGCCGGGTTGTCGCCCTTTTCTACTTTGCCAGCAAACTGGTGCTCAAGGT
5              TCAACTGGGGCCGGGTTGTCGCCCTTTTCTACTTTGCCAGCAAACTGGTGCTCAAGGT
6              TCAACTGGGGCCGGGTTGTCGCCCTTTTCTACTTTGCCAGCAAACTGGTGCTCAAGGT
7              TCAACTGGGGCCGGGTTGTCGCCCTTTTCTACTTTGCCAGCAAACTGGTGCTCAAGGT
8              TCAACTGGGGCCGGGTTGTCGCCCTTTTCTACTTTGCCAGCAAACTGGTGCTCAAGGT
9              TCAACTGGGGCCGGGTTGTCGCCCTTTTCTACTTTGCCAGCAAACTGGTGCTCAAGGT
               ************************************************************

1              GGGCAGCTGCAGGGCAGTGAGCCCAGGGATG 146
2              GGGCAGCTGCAGGGCAGTGAGCCCAGGGATG
3              GGGCAGCTGCAGGGCAGTGAGCCCAGGGATG
4              GGGCAGCTGCAGGGCAGTGAGCCCAGGGATG
5              GGGCAGCTGCAGGGCAGTGAGCCCAGGGATG
6              GGGCAGCTGCAGGGCAGTGAGCCCAGGGATG
7              GGGCAGCTGCAGGGCAGTGAGCCCAGGGATG
8              GGGCAGCTGCAGGGCAGTGAGCCCAGGGATG
9              GGGCAGCTGCAGGGCAGTGAGCCCAGGGATG
               ********************************
```

The database exonic sequence is shown in red (from position 13 to 131) and is 119bp in size. The size of the PCR product is 146bp, which includes the flanking intronic sequence and the entire exonic sequence.

1 = sequence of the exon from the database (http://www.ncbi.nlm.nih/nucleotide)

2 = forward sequence
3 = reverse complement sequence of the reverse } an unaffected individual

4 = forward sequence
5 = reverse complement sequence of the reverse } affected individual from pedigree 1

6 = forward sequence
7 = reverse complement sequence of the reverse } affected individual from pedigree 2

8 = forward sequence
9 = reverse complement sequence of the reverse } affected individual from pedigree 5

Table 15 CLUSTALW 1.81 analysis of exon 5 of *BAX*.

```
1                                                    TAACGCCCACTCCACTCCCCACAGGCCCT 29
2                                                    TAACGCCCACTCCACTCCCCACAGGCCCT
3                                                    TAACGCCCACTCCACTCCCCACAGGCCCT
4                                                    TAACGCCCACTCCACTCCCCACAGGCCCT
5                                                    TAACGCCCACTCCACTCCCCACAGGCCCT
6                                                    TAACGCCCACTCCACTCCCCACAGGCCCT
7                                                    TAACGCCCACTCCACTCCCCACAGGCCCT
8                                                    TAACGCCCACTCCACTCCCCACAGGCCCT
9                                                    TAACGCCCACTCCACTCCCCACAGGCCCT
                                                     ******************************

1                GTGCACCAAGGTGCCGGAACTGATCAGAACCATCATGGGCTGGACATTGGACTTCCTCC 88
2                GTGCACCAAGGTGCCGGAACTGATCAGAACCATCATGGGCTGGACATTGGACTTCCTCC
3                GTGCACCAAGGTGCCGGAACTGATCAGAACCATCATGGGCTGGACATTGGACTTCCTCC
4                GTGCACCAAGGTGCCGGAACTGATCAGAACCATCATGGGCTGGACATTGGACTTCCTCC
5                GTGCACCAAGGTGCCGGAACTGATCAGAACCATCATGGGCTGGACATTGGACTTCCTCC
6                GTGCACCAAGGTGCCGGAACTGATCAGAACCATCATGGGCTGGACATTGGACTTCCTCC
7                GTGCACCAAGGTGCCGGAACTGATCAGAACCATCATGGGCTGGACATTGGACTTCCTCC
8                GTGCACCAAGGTGCCGGAACTGATCAGAACCATCATGGGCTGGACATTGGACTTCCTCC
9                GTGCACCAAGGTGCCGGAACTGATCAGAACCATCATGGGCTGGACATTGGACTTCCTCC
                 ************************************************************

1                GGGAGCGGCTGTTGGGCTGGATCCAAGACCAGGGTGGTTGGGTGAGA             125
2                GGGAGCGGCTGTTGGGCTGGATCCAAGACCAGGGTGGTTGGGTGAGA
3                GGGAGCGGCTGTTGGGCTGGATCCAAGACCAGGGTGGTTGGGTGAGA
4                GGGAGCGGCTGTTGGGCTGGATCCAAGACCAGGGTGGTTGGGTGAGA
5                GGGAGCGGCTGTTGGGCTGGATCCAAGACCAGGGTGGTTGGGTGAGA
6                GGGAGCGGCTGTTGGGCTGGATCCAAGACCAGGGTGGTTGGGTGAGA
7                GGGAGCGGCTGTTGGGCTGGATCCCAGACCAGGGTGGTTGGGTGAGA
8                GGGAGCGGCTGTTGGGCTGGATCCAAGACCAGGGTGGTTGGGTGAGA
9                GGGAGCGGCTGTTGGGCTGGATCCAAGACCAGGGTGGTTGGGTGAGA
                 ************************************************
```

The database exonic sequence is shown in red (from position 24 to 98) and is 75bp in size. The size of the PCR product is 126bp, which includes the flanking intronic sequence and the entire exonic sequence

1 = sequence of the exon from the database (http://www.ncbi.nlm.nih/nucleotide)

2 = forward sequence
3 = reverse complement sequence of the reverse } an unaffected individual

4 = forward sequence
5 = reverse complement sequence of the reverse } affected individual from pedigree 1

6 = forward sequence
7 = reverse complement sequence of the reverse } affected individual from pedigree 2

8 = forward sequence
9 = reverse complement sequence of the reverse } affected individual from pedigree 5

Table 16 CLUSTALW 1.81 analysis of exon 6 of *BAX*.

```
1                                                       CCCTGTCTCCAGGACGGCCTCCTCTC 26
2                                                       CCCTGTCTCCAGGACGGCCTCCTCTC
4                                                       CCCTGTCTCCAGGACGGCCTCCTCTC
5                                                       CCCTGTCTCCAGGACGGCCTCCTCTC
6                                                       CCCTGTCTCCAGGACGGCCTCCTCTC
7                                                       CCCTGTCTCCAGGACGGCCTCCTCTC
8                                                       CCCTGTCTCCAGGACGGCCTCCTCTC
9                                                       CCCTGTCTCCAGGACGGCCTCCTCTC
                                                        **************************

1               CTACTTTGGGACGCCCACGTGGCAGACCGTGACCATCTTTGTGGCGGGAGTGCTCACCGC 86
2               CTACTTTGGGACGCCCACGTGGCAGACCGTGACCATCTTTGTGGCGGGAGTGCTCACCGC
4               CTACTTTGGGACGCCCACGTGGCAGACCGTGACCATCTTTGTGGCGGGAGTGCTCACCGC
5               CTACTTTGGGACGCCCACGTGGCAGACCGTGACCATCTTTGTGGCGGGAGTGCTCACCGC
6               CTACTTTGGGACGCCCACGTGGCAGACCGTGACCATCTTTGTGGCGGGAGTGCTCACCGC
7               CTACTTTGGGACGCCCACGTGGCAGACCGTGACCATCTTTGTGGCGGGAGTGCTCACCGC
8               CTACTTTGGGACGCCCACGTGGCAGACCGTGACCATCTTTGTGGCGGGAGTGCTCACCGC
9               CTACTTTGGGACGCCCACGTGGCAGACCGTGACCATCTTTGTGGCGGGAGTGCTCACCGC
                ************************************************************

1               CTCGCTCACCATCTGGAAGAAGATGGGCTGAGGCCCCCAGCTGCCTTGGACTGTGTT 143
2               CTCACTCACCATCTGGAAGAAGATGGGCTGAGGCCCCCAGCTGCCTTGGACTGTGTT
4               CTCACTCACCATCTGGAAGAAGATGGGCTGAGGCCCCCAGCTGCCTTGGACTGTGTT
5               CTCACTCACCATCTGGAAGAAGATGGGCTGAGGCCCCCAGCTGCCTTGGACTGTGTT
6               CTCACTCACCATCTGGAAGAAGATGGGCTGAGGCCCCCAGCTGCCTTGGACTGTGTT
7               CTCACTCACCATCTGGAAGAAGATGGGCTGAGGCCCCCAGCTGCCTTGGACTGTGTT
8               CTCACTCACCATCTGGAAGAAGATGGGCTGAGGCCCCCAGCTGCCTTGGACTGTGTT
9               CTCACTCACCATCTGGAAGAAGATGGGCTGAGGCCCCCAGCTGCCTTGGACTGTGTT
                *** ********************************************************

1               TTTCCTCCATAAATTATGGCATT 166
2               TTTCCTCCATAAATTATGGCATT
4               TTTCCTCCATAAATTATGGCATT
5               TTTCCTCCATAAATTATGGCATT
6               TTTCCTCCATAAATTATGGCATT
7               TTTCCTCCATAAATTATGGCATT
8               TTTCCTCCATAAATTATGGCATT
9               TTTTCTCCATAAATTATGGCATT
                ***********************
```

The database exonic sequence is shown in red (from position 13 to 138) and is 126bp in size. The size of the PCR product is 166bp, which includes the flanking intronic sequence and the entire exonic sequence.

An error was found in the submitted database sequence of exon 6 (L22473) of *BAX* at base pair position 552 (base pair position 90 in table 16, shown in blue). The BLAST search results of exon 6 sequence against the human genome database indicated that the G at base pair position 1756471 on the NT_011109 contig should be replaced by an A (see table 16).

1 = sequence of the exon from the database (http://www.ncbi.nlm.nih/nucleotide)

2 = forward sequence
3 = reverse complement sequence of the reverse } an unaffected individual

4 = forward sequence
5 = reverse complement sequence of the reverse } affected individual from pedigree 1

6 = forward sequence
7 = reverse complement sequence of the reverse } affected individual from pedigree 2

8 = forward sequence
9 = reverse complement sequence of the reverse } affected individual from pedigree 5

GLYCOGEN SYNTHASE (*GSY1*)

Table 17 ClustalW 1.81 alignment of sequence of the UTR5A region of *GSY1*

```
1                                                                  CCCCCTTCTCCC 12
2                                                                  CCCCCTTCTCCC
3                                                                  CCCCCTTCTCCC
4                                                                  CCCCCCTCTCCC
5                                                                  CCCCCTTCTCCC
6                                                                  CCCCCTTCTCCC
7                                                                  CCCCCTTCTCCC
8                                                                  CCCCCTTCTCCC
9                                                                  CCCCCTTCTCCC
                                                                   ************

1                    CCAGATCTGGGTGGCCGAGCTTCCTCCTCTTTCAGACCCAGGAGTCTGAACCCCCAG 69
2                    CCAGATCTGGGTGGCCGAGCTTCCTCCTCTTTCAGACCCAGGAGTCTGAACCCCCAG
3                    CCAGATCTGGGTGGCCGAGCTTCCTCCTCTTTCAGACCCAGGAGTCTGAACCCCCAG
4                    CCAGATCTGGGTGGCCGAGCTTCCTCCTCTTTCAGACCCAGGGAGTCTGAACCCCCA
5                    CCAGATCTGGGTGGCCGAGCTTCCTCCTCTTTCAGACCCAGGAGTCTGAACCCCCAG
6                    CCAGATCTGGGTGGCCGAGCTTCCTCCTCTTTCAGACCCAGGAGTCTGAACCCCCAG
7                    CCAGATCTGGGTGGCCGAGCTTCCTCCTCTTTCAGACCCAGGAGTCTGAACCCCCAG
8                    CCAGATCTGGGTGGCCGAGCTTCCTCCTCTTTCAGACCCAGGAGTCTGAACCCCCAG
9                    CCAGATCTGGGTGGCCGAGCTTCCTCCTCTTTCAGACCCAGGAGTCTGAACCCCCAG
                     ************************************************************

1                  ATCCCCCCTCCCTCAGACCCACCAGTCAGGGCCTCCAGCATCCTCCTCTCTCAAACTCG 128
2                  ATCCCCCCTCCCTCAGACCCAGGAGTCAGGGCCTCCAGCATCCTCCTCTCTCAAACTCG
3                  ATCCCCCCTCCCTCAGACCCAGGAGTCAGGGCCTCCAGCATCCTCCTCTCTCAAACTCG
4                  ATCCCCCCTCCCTCAGACCCAGGAGTCAGGGCCTCCAGCATCCTCCTCTCTCAAACTCG
5                  ATCCCCCCTCCCTCAGACCCAGGAGTCAGGGCCTCCAGCATCCTCCTCTCTCAAACTCG
6                  ATCCCCCCTCCCTCAGACCCAGGAGTCAGGGCCTCCAGCATCCTCCTCTCTCAAACTCG
7                  ATCCCCCCTCCCTCAGACCCAGGAGTCAGGGCCTCCAGCATCCTCCTCTCTCAAACTCG
8                  ATCCCCCCTCCCTCAGACCCAGGAGTCAGGGCCTCCAGCATCCTCCTCTCTCAAACTCG
9                  ATCCCCCCTCCCTCAGACCCAGGAGTCAGGGCCTCCAGCATCCTCCTCTCTCAAACTCG
                   ********************  ***************************************

1                AGACTGCTAGCTCGCTCCTCACCCTCCATGCACCCACACTTACAACGGTTGCCATGATGC 188
2                AGACTGCTAGCTCGCTCCTCACCCTCCATGCACCCACACTTACAACGGTTGCCATGATGC
3                AGACTGCTAGCTCGCTCCTCACCCTCCATGCACCCACACTTACAACGGTTGCCATGATGC
4                AGACTGCTAGCTCGCTCCTCACCCTCCATGCACCCACACTTACAACGGTTGCCATGATGC
5                AGACTGCTAGCTCGCTCCTCACCCTCCATGCACCCACACTTACAACGGTTGCCATGATGC
6                AGACTGCTAGCTCGCTCCTCACCCTCCATGCACCCACACTTACAACGGTTGCCATGATGC
7                AGACTGCTAGCTCGCTCCTCACCCTCCATGCACCCACACTTACAACGGTTGCCATGATGC
8                AGACTGCTAGCTCGCTCCTCACCCTCCATGCACCCACACTTACAACGGTTGCCATGATGC
9                AGACTGCTAGCTCGCTCCTCACCCTCCATGCACCCACACTTACAACGGTTGCCATGATGC
                 ************************************************************

1                  TCACCAACTGCCAGAGTCCTAGCGGAAACGGAAACGCGCGAGGAGCGAGATGTTCTGG 246
2                  TCACCAACTGCCAGAGTCCTAGCGGAAACGGAAACGCGCGAGGAGCGAGATCTTCTGG
3                  TCACCAACTGCCAGAGTCCTAGCGGAAACGGAAACGCGCGAGGAGCGAGATCTTCTGG
4                  TCACCAACTGCCAGAGTCCTAGCGGAAACGGAAACGCGCGAGGAGCGAGATCTTCTGG
5                  TCACCAACTGCCAGAGTCCTAGCGGAAACGGAAACGCGCGAGGAGCGAGATCTTCTGG
6                  TCACCAACTGCCAGAGTCCTAGCGGAAACGGAAACGCGCGAGGAGCGAGATCTTCTGG
7                  TCACCAACTGCCAGAGTCCTAGCGGAAACGGAAACGCGCGAGGAGXGAGATCTTCTGG
8                  TCACCAACTGCCAGAGTCCTAGCGGAAACGGAAACGCGCGAGGAGCGAGATCTTCTGG
9                  TCACCAACTGCCAGAGTCCTAGCGGAAACGGAAACGCGCGAGGAGCGAGATCTTCTGG
                   ************************************************************
```

UTR5 region of *GSY1* was subdivided into the UTR5A, UTR5B and UTR5C regions. The UTR5A region extends from 1 to 246bp. The overlapping sequence of the UTR5A region with the UTR5B region is underlined.

In the case of *GSY1*, it was also found that some other interesting results emerged, namely the database sequence of UTR5A (acc no. S79190) at position 139 and 140 was incorrect. The nucleotide CC (shown in red) should be replaced with a GG (table 17 in Appendix III, base pair position 91 and 92). This result was verified by a BLAST homology search of the UTR5A The database sequence against the Human genome database (http://www.ncbi.nlm.nih.gov/human genome).

1 = sequence of the exon from the database (http://www.ncbi.nlm.nih/nucleotide)
2 = forward sequence
3 = reverse complement sequence of the reverse } an unaffected individual
4 = forward sequence
5 = reverse complement sequence of the reverse } affected individual from pedigree 1
6 = forward sequence
7 = reverse complement sequence of the reverse } affected individual from pedigree 2
8 = forward sequence
9 = reverse complement sequence of the reverse } affected individual from pedigree 5

Table 18 ClustalW1.81 alignment of sequence of the UTR5B region of *GSY1*

```
1                                                    TGGTTCCTCATAAATATGTAAATATTGTGGGGGCGT 279
2                                                    TGGTTCCTCATAAATGTGTAAATATTGTGGGGGCGT
3                                                    TGGTTCCTCATAAATATGTAAATATTGTGGGGGCGT
4                                                    TGGTTCCTCATAAATATGTAAATATTGTGGGGGCGT
5                                                    TGGTTCCTCATAAATATGTAAATATTGTGGGGGCGT
7                                                    TGGTTCCTCATAAATATGTAAATATTGTGGGGGCGT
8                                                    TGGTTCCTCATAAATATGTAAATATTGTGGGGGCGT
9                                                    TGGTTCCTCATAAATATGTAAATATTGTGGGGGCGT
                                                     ************************************

1                GGCGGGACTTTATGCCCCGCCCTCCCACCGAGGCTGCGCTGGAAGACGGGCGTCCGG 363
2                GGCGGGACTTTATGCCCCGCCCTCCCACCGAGGCTGCGCTGGAAGACGGGCGTCCGG
3                GGCGGGACTTTATGCCCCGCCCTCCCACCGAGGCTGCGCTGGAAGACGGGCGTCCGG
4                GGCGGGACTTTATGCCCCGCCCTCCCACCGAGGCTGCGCTGGAAGACGGGCGTCCGG
5                GGCGGGACTTTATGCCCCGCCCTCCCACCGAGGCTGCGCTGGAAGACGGGCGTCCGG
7                GGCGGGACTTTATGCCCCGCCCTCCCACCGAGGCTGCGCTGGAAGACGGGCGTCCGG
8                GGCGGGACTTTATGCCCCGCCCTCCCACCGAGGCTGCGCTGGAAGACGGGCGTCCGG
9                GGCGGGACTTTATGCCCCGCCCTCCCACCGAGGCTGCGCTGGAAGACGGGCGTCCGG
                 ************************************************************

1            AACTCCTGGGTCCATTCCCAATGTCTCCTGGCTCCTCCTCCTTTTGCTCAGCCCCACCA 422
2            AACTCCTGGGTCCATTCCCAATGTCTCCTGGCTCCTCCTCCTTTTGCTCAGCCCCACCA
3            AACTCCTGGGTCCATTCCCAATGTCTCCTGGCTCCTCCTCCTTTTGCTCAGCCCCACCA
4            AACTCCTGGGTCCATTCCCAATGTCTCCTGGCTCCTCCTCCTTTTGCTCAGCCCCACCA
5            AACTCCTGGGTCCATTCCCAATGTCTCCTGGCTCCTCCTCCTTTTGCTCAGCCCCACCA
7            AACTCCTGGGTCCATTCCCAATGTCTCCTGGCTCCTCCTCCTTTTGCTCAGCCCCACCA
8            AACTCCTGGGTCCATTCCCAATGTCTCCTGGCTCCTCCTCCTTTTGCTCAGCCCCACCA
9            AACTCCTGGGTCCATTCCCAATGTCTCCTGGCTCCTCCTCCTTTTGCTCAGCCCCACCA
             ************************************************************

1            CCGCCCCGCTCAGGCCGATTGCAAAATTCTTAGCTGTTCTTTGAGTAACCACACGTCCTC 482
2            CCGCCCCGCTCAGGCCGATTGCAAAATTCTTAGCTGTTCTTTGAGTAACCACACGTCCTC
3            CCGCCCCGCTCAGGCCGATTGCAAAATTCTTAGCTGTTCTTTGAGTAACCACACGTCCTC
4            CCGCCCCGCTCAGGCCGATTGCAAAATTCTTAGCTGTTCTTTGAGTAACCACACGTCCTC
5            CCGCCCCGCTCAGGCCGATTGCAAAATTCTTAGCTGTTCTTTGAGTAACCACACGTCCTC
7            CCGCCCCGCTCAGGCCGATTGCAAAATTCTTAGCTGTTCTTTGAGTAACCACACGTCCTC
8            CCGCCCCGCTCAGGCCGATTGCAAAATTCTTAGCTGTTCTTTGAGTAACCACACGTCCTC
9            CCGCCCCGCTCAGGCCGATTGCAAAATTCTTAGCTGTTCTTTGAGTAACCACACGTCCTC
             ************************************************************

1            CACGCTCCTTTCCTCTTTATCCCACCCTCCTCCTCCTCCCCACTTTGAGGCCTCCA 538
2            CACGCTCCTTTCCTCTTTATCCCACCCTCCTCCTCCTCCCCACTTTGAGGCCTCCA
3            CACGCTCCTTTCCTCTTTATCCCACCCTCCTCCTCTTCCCCACTTTGAAGCCTCCA
4            CACGCTCCTTTCCTCTTTATCCCACCCTCCTCCTCCTCCCCACTTTGAGGCCTCCA
5            CACGCTCCTTTCCTCTTTATCCCACCCTCCTCCTCCTCCCCACTTTGAGGCCTCCA
7            CACGCTCCTTTCCTCTTTATCCCACCCTCCTCCTCCTCCCCACTTTGAGGCCTCCA
8            CACGCTCCTTTCCTCTTTATCCCACCCTCCTCCTCCTCCCCACTTTGAGGCCTCCA
9            CACGCTCCTTTCCTCTTTATCCCACCCTCCTCCTGCTCCCCACTTTGAGGCCTCCA
             ************************************************************

1                                                GCTCCAGATTGACAGAGCTGAAGGCC 564
2                                                GCTCCAGATTGACAGAGCTGAAGGCC
3                                                GCTCCAGATTGACAGAGCTGAAGGCC
4                                                GCTCCAGATTGACAGAGCTGAAGGCC
5                                                GCTCCAGATTGACAGAGCTGAAGGCC
7                                                GCTCCAGATTGACAGAGCTGAAGGCC
8                                                GCTCCAGATTGACAGAGCTGAAGGCC
9                                                GCTCCAGATTGACAGAGCTGAAGGCC
                                                 **************************
```

The UTR5B region analysed extends from position 243 to 564. The PCR product is 322bp in size and the overlapping sequences of UTR5B with UTR5A and UTR5C are underlined.

1 = sequence of the exon from the database (http://www.ncbi.nlm.nih/nucleotide)

2 = forward sequence
3 = reverse complement sequence of the reverse } an unaffected individual

4 = forward sequence
5 = reverse complement sequence of the reverse } affected individual from pedigree 1

6 = forward sequence
7 = reverse complement sequence of the reverse } affected individual from pedigree 2

8 = forward sequence
9 = reverse complement sequence of the reverse } affected individual from pedigree 5

Table 19 ClustalW 1.81 alignment of sequence of the UTR5C region of *GSY1*

```
1                                                       GAAGGCCACTCGTTTCCGTAAAGCGCA 584
2                                                       GAAGGCCACTCGTTTCCGTAAAGCGCA
3                                                       GAAGGCCACTCGTTTCCGTAAAGCGCA
4                                                       GAAGGCCACTCGTTTCCGTAAAGCGCA
5                                                       GAAGGCCACTAGTTTCCGTAAAGCGCA
6                                                       GAAGGCCACTAGTTTCCGTAAAGCGCA
7                                                       GAAGGCCACTCGTTTCCGTAAAGCGCA
8                                                       GAAGGCCACTCGTTTCCGTAAAGCGCA
9                                                       GAAGGCCACTCGTTTCCGTAAAGCGCA
                                                        ****************************

1                GCGGCAAGCTCCGCCCCCTGGAGGTCTCACGGCCGAGGCGGGGCCGGGACACAGCG 640
2                GCGGCAAGCTCCGCCCCCTGGAGGTCTCACGGCCGAGGCGGGGCCGGGACACAGCG
3                GCGGCAAGCTCCGCCCCCTGGAGGTCTCACGGCCGAGGCGGGGCCGGGACACAGCG
4                GCGGCAAGCTCCGCCCCCTGGAGGTCTCACGGCCGAGGCGGGGCCGGGACACAGCG
5                GCGGCAAGCTCCGCCCCCTGGAGGTCTCACGGCCGAGGCGGGGCCGGGACACAGCG
6                GCGGCAAGCTCCGCCCCCTGGAGGTCTCACGGCCGAGGCGGGGCCGGGACACAGCG
7                GCGGCAAGCTCCGCCCCCTGGAGGTCTCACGGCCGAGGCGGGGCCGGGACACAGCG
8                GCGGCAAGCTCCGCCCCCTGGAGGTCTCACGGCCGAGGCGGGGCCGGGACACAGCG
9                GCGGCAAGCTCCGCCCCCTGGAGGTCTCACGGCCGAGGCGGGGCCGGGACACAGCG
                 ************************************************************

1             ACTCCAACCCGTTCCACCAATGAGGACACGCGGAAATCCCCGCCCCCGGGCGTGGGCCAA 700
2             ACTCCAACCCGTTCCACCAATGAGGACACGCGGAAATCCCCGCCCCCGGGCGTGGGCCAA
3             ACTCCAACCCGTTCCACCAATGAGGACACGCGGAAATCCCCGCCCCCGGGCGTGGGCCAA
4             ACTCCAACCCGTTCCACCAATGAGGACACGCGGAAATCCCCGCCCCCGGGCGTGGGCCAA
5             ACTCCAACCCGTTCCACCAATGAGGACACGCGGAAATCCCCGCCCCCGGGCGTGGGCCAA
6             ACTCCAACCCGTTCCACCAATGAGGACACGCGGAAATCCCCGCCCCCGGGCGTGGGCCAA
7             ACTCCAACCCGTTCCACCAATGAGGACACGCGGAAATCCCCGCCCCCGGGCGTGGGCCAA
8             ACTCCAACCCGTTCCACCAATGAGGACACGCGGAAATCCCCGCCCCCGGGCGTGGGCCAA
9             ACTCCAACCCGTTCCACCAATGAGGACACGCGGAAATCCCCGCCCCCGGGCGTGGGCCAA
              ****************************************************************

1             TGGCGCTGGCTGGGGGGGGCGGCCACCGCCTCCTGGCGGCTGCGAGGTTTCACTGCAGG 760
2             TGGCGCTGGCTGGGGGGGGCGGCCACCGCCTCCTGGCGGCTGCGAGGTTTCACTGCAGG
3             TGGCGCTGGCTGGGGGGGGCGGCCACCGCCTCCTGGCGGCTGCGAGGTTTCACTGCAGG
4             TGGCGCTGGCTGGGGGGGGCGGCCACCGCCTCCTGGCGGCTGCGAGGTTTCACTGCAGG
5             TGGCGCTGGCTGGGGGGGGCGGCCACCGCCTCCTGGCGGCTGCGAGGTTTCACTGCAGG
6             TGGCGCTGGCTGGGGGGGGCGGCCACCGCCTCCTGGCGGCTGCGAGGTTTCACTGCAGG
7             TGGCGCTGGCTGGGGGGGGCGGCCACCGCCTCCTGGCGGCTGCGAGGTTTCACTGCAGG
8             TGGCGCTGGCTGGGGGGGGCGGCCACCGCCTCCTGGCGGCTGCGAGGTTTCACTGCAGG
9             TGGCGCTGGCTGGGGGGGGCGGCCACCGCCTCCTGGCGGCTGCGAGGTTTCACTGCAGG
              ****************************************************************

1             GGCGCCAGTGGGCTCAGTGACGCTGCGGCCTCCTTCTGCCTAGGTCCCAACGCTTCGGGG 820
2             GGCGCCAGTGGGCTCAGTGACGCTGCGGCCTCCTTCTGCCTAGGTCCCAACGCTTCGGGG
3             GGCGCCAGTGGGCTCAGTGACGCTGCGGCCTCCTTCTGCCTAGGTCCCAACGCTTCGGGG
4             GGCGCCAGTGGGCTCAGTGACGCTGCGGCCTCCTTCTGCCTAGGTCCCAACGCTTCGGGG
5             GGCGCCAGTGGGCTCAGTGACGCTGCGGCCTCCTTCTGCCTAGGTCCCAACGCTTCGGGG
6             GGCGCCAGTGGGCTCAGTGACGCTGCGGCCTCCTTCTGCCTAGGTCCCAACGCTTCGGGG
7             GGCGCCAGTGGGCTCAGTGACGCTGCGGCCTCCTTCTGCCTAGGTCCCAACGCTTCGGGG
8             GGCGCCAGTGGGCTCAGTGACGCTGCGGCCTCCTTCTGCCTAGGTCCCAACGCTTCGGGG
9             GGCGCCAGTGGGCTCAGTGACGCTGCGGCCTCCTTCTGCCTAGGTCCCAACGCTTCGGGG
              ****************************************************************
```

```
1          CAGGGGTGCGGTCTTGCAATAGGAAGCCGAGCGTCTTGCAAGCTTCCCGTCGGGCACCA 879
2          CAGGGGTGCGGTCTTGCAATAGGAAGCCGAGCGTCTTGCAAGCTTCCCGTCGGGCACCA
3          CAGGGGTGCGGTCTTGCAATAGGAAGCCGAGCGTCTTGCAAGCTTCCCGTCGGGCACCA
4          CAGGGGTGCGGTCTTGCAATAGGAAGCCGAGCGTCTTGCAAGCTTCCCGTCGGGCACCA
5          CAGGGGTGCGGTCTTGCAATAGGAAGCCGAGCGTCTTGCAAGCTTCCCGTCGGGCACCA
6          CAGGGGTGCGGTCTTGCAATAGGAAGCCGAGCGTCTTGCAAGCTTCCCGTCGGGCACCA
7          CAGGGGTGCGGTCTTGCAATAGGAAGCCGAGCGTCTTGCAAGCTTCCCGTCGGGCACCA
8          CAGGGGTGCGGTCTTGCAATAGGAAGCCGAGCGTCTTGCAAGCTTCCCGTCGGGCACCA
9          CAGGGGTGCGGTCTTGCAATAGGAAGCCGAGCGTCTTGCAAGCTTCCCGTCGGGCACCA
           ************************************************************
```

The UTR5C region analysed extends from position 558 to 879 and the size of the PCR product is 322bp. The overlapping sequences of the UTR5B region with the UTR5C region are underlined.

1 = sequence of the exon from the database (http://www.ncbi.nlm.nih/nucleotide)

2 = forward sequence
3 = reverse complement sequence of the reverse } an unaffected individual

4 = forward sequence
5 = reverse complement sequence of the reverse } affected individual from pedigree 1

6 = forward sequence
7 = reverse complement sequence of the reverse } affected individual from pedigree 2

8 = forward sequence
9 = reverse complement sequence of the reverse } affected individual from pedigree 5

Table 20 ClustalW 1.81 alignment of the exon 1 sequences of *GSY1*

```
1                                                  TGTCTTGCAATAGGAAGCCGAGCGTCTTGCAAGCTTCC 38
2                                                  TGTCTTGCAATAGGAAGCCGAGCGTCTTGCAAGCTTCC
3                                                  TGTCTTGCAATAGGAAGCCGAGCGTCTTGCAAGCTTCC
4                                                  TGTCTTGCAATAGGAAGCCGAGCGTCTTGCAAGCTTCC
5                                                  TGTCTTGCAATAGGAAGCCGAGCGTCTTGCAAGCTTCC
6                                                  TGTCTTGCAATAGGAAGCCGAGCGTCTTGCAAGCTTCC
7                                                  TGTCTTGCAATAGGAAGCCGAGCGTCTTGCAAGCTTCC
8                                                  TGTCTTGCAATAGGAAGCCGAGCGTCTTGCAAGCTTCC
9                                                  TGTCTTGCAATAGGAAGCCGAGCGTCTTGCAAGCTTCC
                                                   ***************************************

1                          CGTCGGGCACCAGCTACTCGGCCCCGCACCCTACCTGGTGCATTCCCTAGACAC 92
2                          CGTCGGGCACCAGCTACTCGGCCCCGCACCCTACCTGGTGCATTCCCTAGACAC
3                          CGTCGGGCACCAGCTACTCGGCCCCGCACCCTACCTGGTGCATTCCCTAGACAC
4                          CGTCGGGCACCAGCTACTCGGCCCCGCACCCTACCTGGTGCATTCCCTAGACAC
5                          CGTCGGGCACCAGCTACTCGGCCCCGCACCCTACCTGGTGCATTCCCTAGACAC
6                          CGTCGGGCACCAGCTACTCGGCCCCGCACCCTACCTGGTGCATTCCCTAGACAC
7                          CGTCGGGCACCAGCTACTCGGCCCCGCACCCTACCTGGTGCATTCCCTAGACAC
8                          CGTCGGGCACCAGCTACTCGGCCCCGCACCCTACCTGGTGCATTCCCTAGACAC
9                          CGTCGGGCACCAGCTACTCGGCCCCGCACCCTACCTGGTGCATTCCCTAGACAC
                           ************************************************************

1                        CTCCGGGGTCCCTACCTGGAGATCCCCGGAGCCCCCCTTCCTGCGCCAGCCATGCC 148
2                        CTCCGGGGTCCCTACCTGGAGATCCCCGGAGCCCCCCTTCCTGCGCCAGCCATGCC
3                        CTCCGGGGTCCCTACCTGGAGATCCCCGGAGCCCCCCTTCCTGCGCCAGCCATGCC
4                        CTCCGGGGTCCCTACCTGGAGATCCCCGGAGCCCCCCTTCCTGCGCCAGCCATGCC
5                        CTCCGGGGTCCCTACCTGGAGATCCCCGGAGCCCCCCTTCCTGCGCCAGCCATGCC
6                        CTCCGGGGTCCCTACCTGGAGATCCCCGGAGCCCCCCTTCCTGCGCCAGCCATGCC
7                        CTCCGGGGTCCCTACCTGGAGATCCCCGGAGCCCCCCTTCCTGCGCCAGCCATGCC
8                        CTCCGGGGTCCCTACCTGGAGATCCCCGAAGCCCCCCTTCCTGCGCCAGCCATGCC
9                        CTCCGGGGTCCCTACCTGGAGATCCCCGGAGCCCCCCTTCCTGCGCCAGCCATGCC
                         ************************************************************

1                      TTTAAACCGCACTTTGTCCATGTCCTCACTGCCAGGACTGGAGGACTGGGAGGATGAA 206
2                      TTTAAACCGCACTTTGTCCATGTCCTCACTGCCAGGACTGGAGGACTGGGAGGATGAA
3                      TTTAAACCGCACTTTGTCCATGTCCTCACTGCCAGGACTGGAGGACTGGGAGGATGAA
4                      TTTAAACCGCACTTTGTCCATGTCCTCACTGCCAGGACTGGAGGACTGGGAGGATGAA
5                      TTTAAACCGCACTTTGTCCATGTCCTCACTGCCAGGACTGGAGGACTGGGAGGATGAA
6                      TTTAAACCGCACTTTGTCCATGTCCTCACTGCCAGGACTGGAGGACTGGGAGGATGAA
7                      TTTAAACCGCACTTTGTCCATGTCCTCACTGCCAGGACTGGAGGACTGGGAGGATGAA
8                      TTTAAACCGCACTTTGTCCATGTCCTCACTGCCAGGACTGGAGGACTGGGAGGATGAA
9                      TTTAAACCGCACTTTGTCCATGTCCTCACTGCCAGGACTGGAGGACTGGGAGGATGAA
                       ************************************************************

1                       TTCGACCTGGAGAACGCAGTGCTCTTCGAAGTGGCCTGGGAGGTGGCTAACAAGGGT 263
2                       TTCGACCTGGAGAACGCAGTGCTCTTCGAAGTGGCCTGGGAGGTGGCTAACAAGGGT
3                       TTCGACCTGGAGAACGCAGTGCTCTTCGAAGTGGCCTGGGAGGTGGCTAACAAGGGT
4                       TTCGACCTGGAGAACGCAGTGCTCTTCGAAGTGGCCTGGGAGGTGGCTAACAAGGGT
5                       TTCGACCTGGAGAACGCAGTGCTCTTCGAAGTGGCCTGGGAGGTGGCTAACAAGGGT
6                       TTCGACCTGGAGAACGCAGTGCTCTTCGAAGTGGCCTGGGAGGTGGCTAACAAGGGT
7                       TTCGACCTGGAGAACGCAGTGCTCTTCGAAGTGGCCTGGGAGGTGGCTAACAAGGGT
8                       TTCGACCTGGAGAACGCAGTGCTCTTCGAAGTGGCCTGGGAGGTGGCTAACAAGGGT
9                       TTCGACCTGGAGAACGCAGTGCTCTTCGAAGTGGCCTGGGAGGTGGCTAACAAGGGT
                        ************************************************************
```

```
1                GAGCACGCCAGGCGT 278
2                GAGCACGCCAGGCGT 278
3                GAGCACGCCAGGCGT
4                GAGCACGCCAGGCGT
5                GAGCACGCCAGGCGT
6                GAGCACGCCAGGCGT
7                GAGCACGCCAGGCGT
8                GAGCACGCCAGGCGT
9                GAGCACGCCAGGCGT
                 ***************
```

The database exonic sequence is shown in red (position 18 to 236) and is 119bp in size. The size of the PCR product is 278bp, which includes the exonic and the flanking intronic regions.

1 = sequence of the exon from the database (http://www.ncbi.nlm.nih/nucleotide)

2 = forward sequence
3 = reverse complement sequence of the reverse } an unaffected individual

4 = forward sequence
5 = reverse complement sequence of the reverse } affected individual from pedigree 1

6 = forward sequence
7 = reverse complement sequence of the reverse } affected individual from pedigree 2

8 = forward sequence
9 = reverse complement sequence of the reverse } affected individual from pedigree 5

Table 21 ClustalW 1.81 alignment of exon 2 sequences of *GSY1*

```
1                                                   CCTGCCCTCACACGCTTGGGCCCACAGTGGGTG 33
2                                                   CCTGCCCTCACACGCTTGGGCCCACAGTGGGTG
3                                                   CCTGCCCTCACACGCTTGGGCCCACAGTGGGTG
4                                                   CCTGCCCTCACACGCTTGGGCCCACAGTGGGTG
5                                                   CCTGCCCTCACACGCTTGGGCCCACAGTGGGTG
6                                                   CCTGCCCTCACACGCTTGGGCCCACAGTGGGTG
7                                                   CCTGCCCTCACACGCTTGGGCCCACAGTGGGTG
                                                    **********************************

1               GCATCTACACGGTGCTGCAGACGAAGGCGAAGGTGACAGGGGACGAATGGGGCGACAACT 93
2               GCATCTACACGGTGCTGCAGACGAAGGCGAAGGTGACAGGGGACGAATGGGGCGACAACT
3               GCATCTACACGGTGCTGCAGACGAAGGCGAAGGTGACAGGGGACGAATGGGGCGACAACT
4               GCATCTACACGGTGCTGCAGACGAAGGCGAAGGTGACAGGGGACGAATGGGGCGACAACT
5               GCATCTACACGGTGCTGCAGACGAAGGCGAAGGTGACAGGGGACGAATGGGGCGACAACT
6               GCATCTACACGGTGCTGCAGACGAAGGCGAAGGTGACAGGGGACGAATGGGGCGACAACT
7               GCATCTACACGGTGCTGCAGACGAAGGCGAAGGTGACAGGGGACGAATGGGGCGACAACT
                ************************************************************

1               ACTTCCTGGTGGGGCCGTACACGGAGCAGGGCGTGAGGACCCAGGTGGAACTGCTGGAGG 153
2               ACTTCCTGGTGGGGCCGTACACGGAGCAGGGCGTGAGGACCCAGGTGGAACTGCTGGAGG
3               ACTTCCTGGTGGGGCCGTACACGGAGCAGGGCGTGAGGACCCAGGTGGAACTGCTGGAGG
4               ACTTCCTGGTGGGGCCGTACACGGAGCAGGGCGTGAGGACCCAGGTGGAACTGCTGGAGG
5               ACTTCCTGGTGGGGCCGTACACGGAGCAGGGCGTGAGGACCCAGGTGGAACTGCTGGAGG
6               ACTTCCTGGTGGGGCCGTACACGGAGCAGGGCGTGAGGACCCAGGTGGAACTGCTGGAGG
7               ACTTCCTGGTGGGGCCGTACACGGAGCAGGGCGTGAGGACCCAGGTGGAACTGCTGGAGG
                ************************************************************

1                 CCCCCACCCCGGCCCTGAAGAGGACACTGGATTCCATGAACAGCAAGGGCTGCAAGGT 211
2                 CCCCCACCCCGGCCCTGAAGAGGACACTGGATTCCATGAACAGCAAGGGCTGCAAGGT
3                 CCCCCACCCCGGCCCTGAAGAGGACACTGGATTCCATGAACAGCAAGGGCTGCAAGGT
4                 CCCCCACCCCGGCCCTGAAGAGGACACTGGATTCCATGAACAGCAAGGGCTGCAAGGT
5                 CCCCCACCCCGGCCCTGAAGAGGACACTGGATTCCATGAACAGCAAGGGCTGCAAGGT
6                 CCCCCACCCCGGCCCTGAAGAGGACACTGGATTCCATGAACAGCAAGGGCTGCAAGGT
7                 CCCCCACCCCGGCCCTGAAGAGGACACTGGATTCCATGAACAGCAAGGGCTGCAAGGT
                  ************************************************************

1               GGGACGTGGCCCAGCCCAGGGCAGAAGCGGGTGGACAGCCTGCA 255
2               GGGACGTGGCCCAGCCCAGGGCAGAAGCGGGTGGACAGCCTGCA
3               GGGACGTGGCCCAGCCCAGGGCAGAAGCGGGTGGACAGCCTGCA
4               GGGACGTGGCCCAGCCCAGGGCAGAAGCGGGTGGACAGCCTGCA
5               GGGACGTGGCCCAGCCCAGGGCAGAAGCGGGTGGACAGCCTGCA
6               GGGACGTGGCCCAGCCCAGGGCAGAAGCGGGTGGACAGCCTGCA
7               GGGACGTGGCCCAGCCCAGGGCAGAAGCGGGTGGACAGCCTGCA
                8********************************************
```

The database exonic sequence is shown in red (from position 28 to 209) and is 172bp in size. The size of the PCR product is 255bp, which includes the exonic and the flanking intronic regions.

1 = sequence of the exon from the database (http://www.ncbi.nlm.nih/nucleotide)

2 = forward sequence
3 = reverse complement sequence of the reverse } an unaffected individual

4 = forward sequence
5 = reverse complement sequence of the reverse } affected individual from pedigree 1

6 = forward sequence
7 = reverse complement sequence of the reverse } affected individual from pedigree 2

8 = forward sequence
9 = reverse complement sequence of the reverse } affected individual from pedigree 5

Table 22 ClustalW 1.81 alignment of the exon 3 sequences of *GSY1*

```
1                                                CAACCCCATCCCAGGTGTATTTCGGGCGCTGGCTGAT 37
2                                                CAACCCCATCCCAGGTGTATTTCGGGCGCTGGCTGAT
3                                                CAACCCCATCCCAGGTGTATTTCGGGCGCTGGCTGAT
4                                                CAACCCCATCCCAGGTGTATTTCGGGCGCTGGCTGAT
5                                                CAACCCCATCCCAGGTGTATTTCGGGCGCTGGCTGAT
6                                                CAACCCCATCCCAGGTGTATTTCGGGCGCTGGCTGAT
7                                                CAACCCCATCCCAGGTGTATTTCGGGCGCTGGCTGAT
8                                                CAACCCCATCCCAGGTGTATTTCGGGCGCTGGCTGAT
9                                                CAACCCCATCCCAGGTGTATTTCGGGCGCTGGCTGAT
                                                 ************************************

1                CGAGGGAGGCCCTCTGGTGGTGCTCCTGGACGTGGGTGCCTCAGCTTGGGCCCTGGAGC 96
2                CGAGGGAGGCCCTCTGGTGGTGCTCCTGGACGTGGGTGCCTCAGCTTGGGCCCTGGAGC
3                CGAGGGAGGCCCTCTGGTGGTGCTCCTGGACGTGGGTGCCTCAGCTTGGGCCCTGGAGC
4                CGAGGGAGGCCCTCTGGTGGTGCTCCTGGACGTGGGTGCCTCAGCTTGGGCCCTGGAGC
5                CGAGGGAGGCCCTCTGGTGGTGCTCCTGGACGTGGGTGCCTCAGCTTGGGCCCTGGAGC
6                CGAGGGAGGCCCTCTGGTGGTGCTCCTGGACGTGGGTGCCTCAGCTTGGGCCCTGGAGC
7                CGAGGGAGGCCCTCTGGTGGTGCTCCTGGACGTGGGTGCCTCAGCTTGGGCCCTGGAGC
8                CGAGGGAGGCCCTCTGGTGGTGCTCCTGGACGTGGGTGCCTCAGCTTGGGCCCTGGAGC
9                CGAGGGAGGCCCTCTGGTGGTGCTCCTGGACGTGGGTGCCTCAGCTTGGGCCCTGGAGC
                 ************************************************************

1                GCTGGAAGGGAGAGCTCTGGGATACCTGCAACATCGGAGTGCCGTGGTACGACCGCGAGG 156
2                GCTGGAAGGGAGAGCTCTGGGATACCTGCAACATCGGAGTGCCGTGGTACGACCGCGAGG
3                GCTGGAAGGGAGAGCTCTGGGATACCTGCAACATCGGAGTGCCGTGGTACGACCGCGAGG
4                GCTGGAAGGGAGAGCTCTGGGATACCTGCAACATCGGAGTGCCGTGGTACGACCGCGAGG
5                GCTGGAAGGGAGAGCTCTGGGATACCTGCAACATCGGAGTGCCGTGGTACGACCGCGAGG
6                GCTGGAAGGGAGAGCTCTGGGATACCTGCAACATCGGAGTGCCGTGGTACGACCGCGAGG
7                GCTGGAAGGGAGAGCTCTGGGATACCTGCAACATCGGAGTGCCGTGGTACGACCGCGAGG
8                GCTGGAAGGGAGAGCTCTGG-ATACCTGCAACATCGGAGTGCCGTGGTACGACCGCGAGG
9                GCTGGAAGGGAGAGCTCTGGGATACCTGCAACATCGGAGTGCCGTGGTACGACCGCGAGG
                 ******************** ****************************************

1                CCAACGACGCTGTCCTCTTTGGCTTTCTGACCACCTGGTTCCTGGGTGAGGTAGGCCCCG 216
2                CCAACGACGCTGTCCTCTTTGGCTTTCTGACCACCTGGTTCCTGGGTGAGGTAGGCCCCG
3                CCAACGACGCTGTCCTCTTTGGCTTTCTGACCACCTGGTTCCTGGGTGAGGTAGGCCCCG
4                CCAACGACGCTGTCCTCTTTGGCTTTCTGACCACCTGGTTCCTGGGTGAGGTAGGCCCCG
5                CCAACGACGCTGTCCTCTTTGGCTTTCTGACCACCTGGTTCCTGGGTGAGGTAGGCCCCG
6                CCAACGACGCTGTCCTCTTTGGCTTTCTGACCACCTGGTTCCTGGGTGAGGTAGGCCCCG
7                CCAACGACGCTGTCCTCTTTGGCTTTCTGACCACCTGGTTCCTGGGTGAGGTAGGCCCCG
8                CCAACGACGCTGTCCTCTTTGGCTTTCTGACCACCTGGTTCCTGGGTGAGGTAGGCCCCG
9                CCAACGACGCTGTCCTCTTTGGCTTTCTGACCACCTGGTTCCTGGGTGAGGTAGGCCCCG
                 ************************************************************
```

The database exonic sequence is shown in red (from position 15 to 167) and is 153bp in size. The size of the PCR product is 216bp, which includes the exonic and the flanking intronic regions.

1 = sequence of the exon from the database (http.//www.ncbi.nlm.nih/nucleotide)

2 = forward sequence 3 = reverse complement sequence of the reverse	an unaffected individual
4 = forward sequence 5 = reverse complement sequence of the reverse	affected individual from pedigree 1
6 = forward sequence 7 = reverse complement sequence of the reverse	affected individual from pedigree 2
8 = forward sequence 9 = reverse complement sequence of the reverse	affected individual from pedigree 5

Table 23 ClustalW 1.81 alignment of the exon 6 sequences of *GSY1*

```
1                CATGGGACCCTCACCCCCAATCCCAGATATTGTGACCCCCAATGGGCTGAATGTG 55
2                CATGGGACCCTCACCCCCAATCCCAGATATTGTGACCCCCAATGGGCTGAATGTG
3                CATGGGACCCTCACCCCCAATCCCAGATATTGTGACCCCCAATGGGCTGAATGTG
4                CATGGGACCCTCACCCCCAATCCCAGATATTGTGACCCCCAATGGGCTGAATGTG
5                CATGGGACCCTCACCCCCAATCCCAGATATTGTGACCCCCAATGGGCTGAATGTG
6                CATGGGACCCTCACCCCCAATCCCAGATATTGTGACCCCCAATGGGCTGAATGTG
7                CATGGGACCCTCACCCCCAATCCCAGATATTGTGACCCCCAATGGGCTGAATGTG
8                CATGGGACCCTCACCCCCAATCCCAGATATTGTGACCCCCAATGGGCTGAATGTG
9                CATGGGACCCTCACCCCCAATCCCAGATATTGTGACCCCCAATGGGCTGAATGTG
                 ************************************************************

1                AAGAAGTTTTCTGCCATGCATGAGTTCCAGAACCTCCATGCTCAGAGCAAGGCTCGAATC 115
2                AAGAAGTTTTCTGCCATGCATGAGTTCCAGAACCTCCATGCTCAGAGCAAGGCTCGAATC
3                AAGAAGTTTTCTGCCATGCATGAGTTCCAGAACCTCCATGCTCAGAGCAAGGCTCGAATC
4                AAGAAGTTTTCTGCCATGCATGAGTTCCAGAACCTCCATGCTCAGAGCAAGGCTCGAATC
5                AAGAAGTTTTCTGCCATGCATGAGTTCCAGAACCTCCATGCTCAGAGCAAGGCTCGAATC
6                AAGAAGTTTTCTGCCATGCATGAGTTCCAGAACCTCCATGCTCAGAGCAAGGCTCGAATC
7                AAGAAGTTTTCTGCCATGCATGAGTTCCAGAACCTCCATGCTCAGAGCAAGGCTCGAATC
8                AAGAAGTTTTCTGCCATGCATGAGTTCCAGAACCTCCATGCTCAGAGCAAGGCTCGAATC
9                AAGAAGTTTTCTGCCATGCATGAGTTCCAGAACCTCCATGCTCAGAGCAAGGCTCGAATC
                 ************************************************************

1                CAGGAGTTTGTGCGGGGCCATTTTTATGGGTACGTGGGGCATAT 159
2                CAGGAGTTTGTGCGGGGCCATTTTTATGGGTACGTGGGGCATAT
3                CAGGAGTTTGTGCGGGGCCATTTTTATGGGTACGTGGGGCATAT
4                CAGGAGTTTGTGCGGGGCCATTTTTATGGGTACGTGGGGCATAT
5                CAGGAGTTTGTGCGGGGCCATTTTTATGGGTACGTGGGGCATAT
6                CAGGAGTTTGTGCGGGGCCATTTTTATGGGTACGTGGGGCATAT
7                CAGGAGTTTGTGCGGGGCCATTTTTATGGGTAGGTGGGGCATAT
8                CAGGAGTTTGTGCGGGGCCATTTTTATGGGTACGTGGGGCATAT
9                CAGGAGTTTGTGCGGGGCCATTTTTATGGGTACGTGGGGCATAT
                 ********************************************
```

The database exonic sequence is shown in red (from position 27 to 144) and is 118bp in size. The size of the PCR product is 159bp, which includes the exonic and the flanking intronic regions.

1 = sequence of the exon from the database (http://www.ncbi.nlm.nih/nucleotide)

2 = forward sequence
3 = reverse complement sequence of the reverse } an unaffected individual

4 = forward sequence
5 = reverse complement sequence of the reverse } affected individual from pedigree 1

6 = forward sequence
7 = reverse complement sequence of the reverse } affected individual from pedigree 2

8 = forward sequence
9 = reverse complement sequence of the reverse } affected individual from pedigree 5

Table 24 ClustalW 1.81 alignment of the exon 7 sequences of *GSY1*

```
1                                                                GTTCTTTAGGCATC 14
2                                                                GTTCTTTAGCCATC
3                                                                GTTCTTTAGGCATC
4                                                                GTTCTTTAGGCATC
5                                                                GTTCTTTAGGCATC
6                                                                GTTCTTTAGGCATC
7                                                                GTTCTTTAGGCATC
8                                                                GTTCTTTAGGCATC
9                                                                GTTCTTTAGGCATC
                                                                 **************

1              TGGACTTCAACTTGGACAAGACCTTATACTTCTTTATCGCCGGCCGCTATGAGTTCT 71
2              TGGACTTCAACTTGGACAAGACCTTATACTTCTTTATCGCCGGCCGCTATGAGTTCT
3              TGGACTTCAACTTGGACAAGACCTTATACTTCTTTATCGCCGGCCGCTATGAGTTCT
4              TGGACTTCAACTTGGACAAGACCTTATACTTCTTTATCGCCGGCCGCTATGAGTTCT
5              TGGACTTCAACTTGGACAAGACCTTATACTTCTTTATCGCCGGCCGCTATGAGTTCT
6              TGGACTTCAACTTGGACAAGACCTTATACTTCTTTATCGCCGGCCGCTATGAGTTCT
7              TGGACTTCAACTTGGACAAGACCTTATACTTCTTTATCGCCGGCCGCTATGAGTTCT
8              TGGACTTCAACTTGGACAAGACCTTATACTTCTTTATCGCCGGCCGCTATGAGTTCT
9              TGGACTTCAACTTGGACAAGACCTTATACTTCTTTATCGCCGGCCGCTATGAGTTCT
               ************************************************************

1              CCAACAAGGGTGCTGACGTCTTCCTGGAGGCATTGGCTCGGCTCAACTATCTGCTC 127
2              CCAACAAGGGTGCTGACGTCTTYCTGGAGGCATTGGCTCGGCTCAACTATCTGCTC
3              CCAACAAGGGTGCTGACGTCTTCCTGGAGGCATTGGCTCGGCTCAACTATCTGCTC
4              CCAACAAGGGTGCTGACGTCTTYCTGGAGGCATTGGCTCGGCTCAACTATCTGCTC
5              CCAACAAGGGTGCTGACGTCTTRCTGGAGGCATTGGCTCGGCTCAACTATCTGCTC
6              CCAACAAGGGTGCTGACGTCTTCCTGGAGGCATTGGCTCGGCTCAACTATCTGCTC
7              CCAACAAGGGTGCTGACGTCTTCCTGGAGGCATTGGCTCGGCTCAACTATCTGCTC
8              CCAACAAGGGTGCTGACGTCTTCCTGGAGGCATTGGCTCGGCTCAACTATCTGCTC
9              CCAACAAGGGTGCTGACGTCTTCCTGGAGGCATTGGCTCGGCTCAACTATCTGCTC
               ********************* ***********************************

1              AGAGTGAGGCCTGGGCT 144
2              AGAGTGAGGCCTGGGCT
3              AGAGTGAGGCCTGGGCT
4              AGAGTGAGGCCTGGGCT
5              AGAGTGAGGCCTGGGCT
6              AGAGTGAGGCCTGGGCT
7              AGAGTGAGGCCTGGGCT
8              AGAGTGAGGCCTGGGCT
9              AGAGTGAGGCCTGGGCT
               *****************
```

1 = sequence of the exon from the database (http://www.ncbi.nlm.nih/nucleotide)

2 = forward sequence
3 = reverse complement sequence of the reverse
} an unaffected individual

4 = forward sequence
5 = reverse complement sequence of the reverse
} affected individual from pedigree 1

6 = forward sequence
7 = reverse complement sequence of the reverse
} affected individual from pedigree 2

8 = forward sequence
9 = reverse complement sequence of the reverse
} affected individual from pedigree 5

The database exonic sequence is shown in red (from position 10 to 130) and is 121bp in size. The size of the PCR product is 144bp, which includes the exonic and the flanking intronic regions.

A reported SNP in exon 7 (rs5464) of *GSY1* was found in the DNA of an unaffected and an affected individual from pedigree 1 (table 24 in appendix III, base pair position 94 and 95). The alleles observed were C/T in the nucleotide position 1186 of the Genbank nucleotide sequence accession number J04501.

Table 25 ClustalW 1.81 alignment of the exon 8 sequences of *GSY1*

```
1                                                                   CTTTCTTTCTGC 12
2                                                                   CTTTCTTTCTGC
3                                                                   CTTTCTTTCTGC
4                                                                   CTTTCTTTCTGC
5                                                                   CTTTCTTTCTGC
6                                                                   CTTTCTTTCTGC
7                                                                   CTTTCTTTCTGC
8                                                                   CTTTCTTTCTGC
9                                                                   CTTTCTTTCTGC
                                                                    ***********

1               GCAGGTGAACGGCAGCGAGCAGACAGTGGTTGCCTTCTTCATCATGCCAGCGCGGACCAA 72
2               GCAGGTGAACGGCAGCGAGCAGACAGTGGTTGCCTTCTTCATCATGCCAGCGCGGACCAA
3               GCAGGTGAACGGCAGCGAGCAGACAGTGGTTGCCTTCTTCATCATGCCAGCGCGGACCAA
4               GCAGGTGAACGGCAGCGAGCAGACAGTGGTTGCCTTCTTCATCATGCCAGCGCGGACCAA
5               GCAGGTGAACGGCAGCGAGCAGACAGTGGTTGCCTTCTTCATCATGCCAGCGCGGACCAA
6               GCAGGTGAACGGCAGCGAGCAGACAGTGGTTGCCTTCTTCATCATGCCAGCGCGGACCAA
7               GCAGGTGAACGGCAGCGAGCAGACAGTGGTTGCCTTCTTCATCATGCCAGCGCGGACCAA
8               GCAGGTGAACGGCAGCGAGCAGACAGTGGTTGCCTTCTTCATCATGCCAGCGCGGACCAA
9               GCAGGTGAACGGCAGCGAGCAGACAGTGGTTGCCTTCTTCATCATGCCAGCGCGGACCAA
                ************************************************************

1               CAATTTCAACGTGGAAACCCTCAAAGGCCAAGCTGTGCGCAAACAGCTT 121
2               CAATTTCAACGTGGAAACCCTCAAAGGCCAAGCTGTGCGCAAACAGCTT
3               CAATTTCAACGTGGAAACCCTCAAAGGCCAAGCTGTGCGCAAACAGCTT
4               CAATTTCAACGTGGAAACCCTCAAAGGCCAAGCTGTGCGCAAACAGCTT
5               CAATTTCAACGTGGAAACCCTCAAAGGCCAAGCTGTGCGCAAACAGCTT
6               CAATTTCAACGTGGAAACCCTCAAAGGCCAAGCTGTGCGCAAACAGCTT
7               CAATTTCAACGTGGAAACCCTCAAAGGCCAAGCTGTGCGCAAACAGCTT
8               CAATTTCAACGTGGAAACCCTCAAAGGCCAAGCTGTGCGCAAACAGCTT
9               CAATTTCAACGTGGAAACCCTCAAAGGCCAAGCTGTGCGCAAACAGCTT
                ***************************************************
```

The database exonic sequence is shown in red (from position 17 to 106) and is 90bp in size. The size of the PCR product is 121bp, which includes the exonic and the flanking intronic regions.

1 = sequence of the exon from the database (http://www.ncbi.nlm.nih/nucleotide)

2 = forward sequence
3 = reverse complement sequence of the reverse } an unaffected individual

4 = forward sequence
5 = reverse complement sequence of the reverse } affected individual from pedigree 1

6 = forward sequence
7 = reverse complement sequence of the reverse } affected individual from pedigree 2

8 = forward sequence
9 = reverse complement sequence of the reverse } affected individual from pedigree 5

Table 26 ClustalW 1.81 alignment of the exon 9 sequences of *GSY1*

```
1                                                                   TGCAGCTGGTGTG 13
2                                                                   TGCAGCTGGTGTG
3                                                                   TGCAGCTGGTGTG
5                                                                   TGCAGCTGGTGTG
6                                                                   TGCAGCTGGTGTG
7                                                                   TGCAGCTGGTGTG
8                                                                   TGCAGCTGGTGTG
9                                                                   TGCAGCTGGTGTG
                                                                    *************

1                TATGTGACCCTGTTGCTCTGCTTCCTCCAGGGACACGGCCAACACGGTGAAGGAAAAGT 72
2                TATGTGACCCTGTTGCTCTGCTTCCTCCAGGGACACGGCCAACACGGTGAAGGAAAAGT
3                TATGTGACCCTGTTGCTCTGCTTCCTCCAGGGACACGGCCAACACGGTGAAGGAAAAGT
5                TATGTGACCCTGTTGCTCTGCTTCCTCCAGGGACACGGCCAACACGGTGAAGGAAAAGT
6                TATGTGACCCTGTTGCTCTGCTTCTTCCAGGGACACGGCCAACACGGTGAAGGAAAAGT
7                TATGTGACCCTGTTGCTCTGCTTCCTCCAGGGACACGGCCAACACGGTGAAGGAAAAGT
8                TATGTGACCCTGTTCCTCTGCTTGCTCCAGGGACACGGCCAACACGGTGAAGGAAAAGT
9                TATGTGACCCTGTTGCTCTGCTTCCTCCAGGGACACGGCCAACACGGTGAAGGAAAAGT
                 ************************************************************

1                TCGGGAGGAAGCTTTATGAATCCTTACTGGTGTGAGTGCCCCACCCTCAGCCCTGTGC 130
2                TCGGGAGGAAGCTTTATGAATCCTTACTGGTGTGAGTGCCCCACCCTCAGCCCTGTGC
3                TCGGGAGGAAGCTTTATGAATCCTTACTGGTGTGAGTGCCCCACCCTCAGCCCTGTGC
5                TCGGGAGGAAGCTTTATGAATCCTTACTGGTGTGAGTGCCCCACCCTCAGCCCTGTGC
6                TCGGGAGGAAGCTTTATGAATCCTTACTGGTGTGAGTGCCCCACCCTCAGCCCTGTGC
7                TCGGGAGGAAGCTTTATGAATCCTTACTGGTGTGAGTGCCCCACCCTCAGCCCTGTGC
8                TCGGGAGGAAGCTTTATGAATCCTTACTGGTGTGAGTGCCCCACCCTCAGCCCTGTGC
9                TCGGGAGGAAGCTTTATGAATCCTTACTGGTGTGAGTGCCCCACCCTCAGCCCTGTGC
                 ************************************************************

1                CCTCCTAGGCCCT 143
2                CCTCCTAGGCCCT
3                CCTCCTAGGCCCT
5                CCTCCTAGGCCCT
6                CCTCCTAGGCCCT
7                CCTCCTAGGCCCT
8                CCTCCTAGGCCCT
9                CCTCCTAGGCCCT
                 *************
```

The database exonic sequence is shown in red (from position 6 to 103) and is 98bp in size. The size of the PCR product is 143bp, which includes the exonic and the flanking intronic regions.

1 = sequence of the exon from the database (http://www.ncbi.nlm.nih/nucleotide)

2 = forward sequence
3 = reverse complement sequence of the reverse } an unaffected individual

4 = forward sequence
5 = reverse complement sequence of the reverse } affected individual from pedigree 1

6 = forward sequence
7 = reverse complement sequence of the reverse } affected individual from pedigree 2

8 = forward sequence
9 = reverse complement sequence of the reverse } affected individual from pedigree 5

Table 27 ClustalW 1.81 alignment of the exon 10 sequences of *GSY1*

```
1                                                       TGCCTTGCAGTGGGAGCCTT 20
2                                                       TGCCTTGCAGTGGGAGCCTT
3                                                       TGCCTTGCAGTGGGAGCCTT
4                                                       TGCCTTGCAGTGGGAGCCTT
5                                                       TGCCTTGCAGTGGGAGCCTT
6                                                       TGCCTTGCAGTGGGAGCCTT
7                                                       TGCCTTGCAGTGGGAGCCTT
8                                                       TGCCTTGCAGTGGGAGCCTT
9                                                       TGCCTTGCAGTGGGAGCCTT
                                                        ********************

1               CCCGACATGAACAAGATGCTGGATAAGGAAGACTTCACTATGATGAAGAGAGCC 74
2               CCCGACATGAACAAGATGCTGGATAAGGAAGACTTCACTATGATGAAGAGAGCC
3               CCCGACATGAACAAGATGCTGGATAAGGAAGACTTCACTATGATGAAGAGAGCC
4               CCCGACATGAACAAGATGCTGGATAAGGAAGACTTCACTATGATGAAGAGAGCC
5               CCCGACATGAACAAGATGCTGGATAAGGAAGACTTCACTATGATGAAGAGAGCC
6               CCCGACATGAACAAGATGCTGGATAAGGAAGACTTCACTATGATGAAGAGAGCC
7               CCCGACATGAACAAGATGCTGGATAAGGAAGACTTCACTATGATGAAGAGAGCC
8               CCCGACATGAACAAGATGCTGGATAAGGAAGACTTCACTATGATGAAGAGAGCC
9               CCCGACATGAACAAGATGCTGGATAAGGAAGACTTCACTATGATGAAGAGAGCC
                ************************************************************

1               ATCTTTGCAACGCAGGTATGGATTGGAC 102
2               ATCTTTGCAACGCAGGTATGGATTGGAC
3               ATCTTTGCAACGCAGGTATGGATTGGAC
4               ATCTTTGCAACGCAGGTATGGATTGGAC
5               ATCTTTGCAACGCAGGTATGGATTGGAC
6               ATCTTTGCAACGCAGGTATGGATTGGAC
7               ATCTTTGCAACGCAGGTATGGATTGGAC
8               ATCTTTGCAACGCAGGTATGGATTGGAC
9               ATCTTTGCAACGCAGGTATGGATTGGAC
                ****************************
```

The database exonic sequence is shown in red (from position 17 to 89) and is 73bp in size. The size of the PCR product is 102bp, which includes the exonic and the flanking intronic regions.

1 = sequence of the exon from the database (http://www.ncbi.nlm.nih/nucleotide)

2 = forward sequence
3 = reverse complement sequence of the reverse } an unaffected individual

4 = forward sequence
5 = reverse complement sequence of the reverse } affected individual from pedigree 1

6 = forward sequence
7 = reverse complement sequence of the reverse } affected individual from pedigree 2

8 = forward sequence
9 = reverse complement sequence of the reverse } affected individual from pedigree 5

Table 28 ClustalW 1.81 alignment of the exon 13 sequences of *GSY1*

```
1                                                       GGGTCTCTCTCCTGCCTTCCCGCAGCT 28
2                                                       GGGTCTCTCTCCTGCCTTCCCGCAGCT
3                                                       GGGTCTCTCTCCTGCCTTCCCGCAGCT
4                                                       GGGTCTCTCTCCTGCCTTCCCGCAGCT
5                                                       GGGTCTCTCTCCTGCCTTCCCGCAGCT
6                                                       GGGTCTCTCTCCTGCCTTCCCGCAGCT
7                                                       GGGTCTCTCTCCTGCCTTCCCGCAGCT
8                                                       GGGTCTCTCTCCTGCCTTCCCGCAGCT
9                                                       GGGTCTCTCTCCTGCCTTCCCGCAGCT
                                                        ***************************

1                GAGTGCACGGTTATGGGAATCCCCAGTATCTCCACCAATCTCTCCGGCTTCGGCTGCTT 87
2                GAGTGCACGGTTATGGGAATCCCCAGTATCTCCACCAATCTCTCCGGCTTCGGCTGCTT
3                GAGTGCACGGTTATGGGAATCCCCAGTATCTCCACCAATCTCTCCGGCTTCGGCTGCTT
4                GAGTGCACGGTTATGGGAATCCCCAGTATCTCCACCAATCTCTCCGGCTTCGGCTGCTT
5                GAGTGCACGGTTATGGGAATCCCCAGTATCTCCACCAATCTCTCCGGCTTCGGCTGCTT
6                GAGTGCACGGTTATGGGAATCCCCAGTATCTCCACCAATCTCTCCGGCTTCGGCTGCTT
7                GAGTGCACGGTTATGGGAATCCCCAGTATCTCCACCAATCTCTCCGGCTTCGGCTGCTT
8                GAGTGCACGGTTATGGGAATCCCCAGTATCTCCACCAATCTCTCCGGCTTCGGCTGCTT
9                GAGTGCACGGTTATGGGAATCCCCAGTATCTCCACCAATCTCTCCGGCTTCGGCTGCTT
                 ************************************************************

1                CATGGAGGAACACATCGCAGACCCCTCAGCTTACGGTCAGCGCCCAGGCCCCTGAGGAT 146
2                CATGGAGGAACACATCGCAGACCCCTCAGCTTACGGTCAGCGCCCAGGCCCCTGAGGAT
3                CATGGAGGAACACATCGCAGACCCCTCAGCTTACGGTCAGCGCCCAGGCCCCTGAGGAT
4                CATGGAGGAACACATCGCAGACCCCTCAGCTTACGGTCAGCGCCCAGGCCCCTGAGGAT
5                CATGGAGGAACACATCGCAGACCCCTCAGCTTACGGTCAGCGCCCAGGCCCCTGAGGAT
6                CATGGAGGAACACATCGCAGACCCCTCAGCTTACGGTCAGCGCCCAGGCCCCTGAGGAT
7                CATGGAGGAACACATCGCAGACCCCTCAGCTTACGGTCAGCGCCCAGGCCCCTGAGGAT
8                CATGGAGGAACACATCGCAGACCCCTCAGCTTACGGTCAGCGCCCAGGCCCCTGAGGAT
9                CATGGAGGAACACATCGCAGACCCCTCAGCTTACGGTCAGCGCCCAGGCCCCTGAGGAT
                 ************************************************************
```

The database exonic sequence is shown in red (from position 26 to 122) and is 97bp in size. The size of the PCR product is 146bp, which includes the exonic and the flanking intronic regions.

1 = sequence of the exon from the database (http://www.ncbi.nlm.nih/nucleotide)

2 = forward sequence
3 = reverse complement sequence of the reverse } an unaffected individual

4 = forward sequence
5 = reverse complement sequence of the reverse } affected individual from pedigree 1

6 = forward sequence
7 = reverse complement sequence of the reverse } affected individual from pedigree 2

8 = forward sequence
9 = reverse complement sequence of the reverse } affected individual from pedigree 5

Table 29 ClustalW 1.81 alignment of the exon 14 sequences of *GSY1*

```
1                                                                   AGGATGTTTTCTGAAC 16
2                                                                   AGGATGTTTTCTGAAC
3                                                                   AGGATGTTTTCTGAAC
4                                                                   AGGATGTTTTCTGAAC
5                                                                   AGGATGTTTTCTGAAC
6                                                                   AGGATGTTTTCTGAAC
7                                                                   AGGATGTTTTCTGAAC
8                                                                   AGGATGTTTTCTGAAC
9                                                                   AGGATGTTTTCTGAAC
                                                                    ****************

1                CTGGGTCCTGGCCTCCACAGGTATCTACATTCTTGACCGGCGGTTCCGCAGCCTGGATG 98
2                CTGGGTCCTGGCCTCCACAGGTATCTACATTCTTGACCGGCGGTTCCGCAGCCTGGATG
3                CTGGGTCCTGGCCTCCACAGGTATCTACATTCTTGACCGGCGGTTCCGCAGCCTGGATG
4                CTGGGTCCTGGCCTCCACAGGTATCTACATTCTTGACCGGCGGTTCCGCAGCCTGGATG
5                CTGGGTCCTGGCCTCCACAGGTATCTACATTCTTGACCGGCGGTTCCGCAGCCTGGATG
6                CTGGGTCCTGGCCTCCACAGGTATCTACATTCTTGACCGGCGGTTCCGCAGCCTGGATG
7                CTGGGTCCTGGCCTCCACAGGTATCTACATTCTTGACCGGCGGTTCCGCAGCCTGGATG
8                CTGGGTCCTGGCCTCCACAGGTATCTACATTCTTGACCGGCGGTTCCGCAGCCTGGATG
9                CTGGGTCCTGGCCTCCACAGGTATCTACATTCTTGACCGGCGGTTCCGCAGCCTGGATG
                 ************************************************************

1                ATTCCTGCTCGCAGCTCACCTCCTTCCTCTACAGTTTCTGTCAGCAGAGCCGGCGGCAG 157
2                ATTCCTGCTCGCAGCTCACCTCCTTCCTCTACAGTTTCTGTCAGCAGAGCCGGCGGCAG
3                ATTCCTGCTCGCAGCTCACCTCCTTCCTCTACAGTTTCTGTCAGCAGAGCCGGCGGCAG
4                ATTCCTGCTCGCAGCTCACCTCCTTCCTCTACAGTTTCTGTCAGCAGAGCCGGCGGCAG
5                ATTCCTGCTCGCAGCTCACCTCCTTCCTCTACAGTTTCTGTCAGCAGAGCCGGCGGCAG
6                ATTCCTGCTCGCAGCTCACCTCCTTCCTCTACAGTTTCTGTCAGCAGAGCCGGCGGCAG
7                ATTCCTGCTCGCAGCTCACCTCCTTCCTCTACAGTTTCTGTCAGCAGAGCCGGCGGCAG
8                ATTCCTGCTCGCAGCTCACCTCCTTCCTCTACAGTTTCTGTCAGCAGAGCCGGCGGCAG
9                ATTCCTGCTCGCAGCTCACCTCCTTCCTCTACAGTTTCTGTCAGCAGAGCCGGCGGCAG
                 ************************************************************

1                 CGTATCATCCAGCGGAACCGCACGGAGCGCCTCTCCGACCTTCTGGACTGGAAATAC 214
2                 CGTATCATCCAGCGGAACCGCACGGAGCGCCTCTCCGACCTTCTGGACTGGAAATAC
3                 CGTATCATCCAGCGGAACCGCACGGAGCGCCTCTCCGACCTTCTGGACTGGAAATAC
4                 CGTATCATCCAGCGGAACCGCACGGAGCGCCTTTCCGACCTTCTGGACTGGAAATAC
5                 CGTATCATCCAGCGGAACCGCACGGAGCGCCTCTCCGACCTTCTGGACTGGAAATAC
6                 CGTATCATCCAGCGGAACCGCACGGAGCGCCTCTCCGACCTTCTGGACTGGAAATAC
7                 CGTATCATCCAGCGGAACCGCACGGAGCGCCTCTCCGACCTTCTGGACTGGAAATAC
8                 CGTATCATCCAGCGGAACCGCACGGAGCGCCTCTCCGACCTTCTGGACTGGAAATAC
9                 CGTATCATCCAGCGGAACCGCACGGAGCGCCTCTCCGACCTTCTGGACTGGAAATAC
                  ************************************************************

1                CTAGGCCGGGTAGGACCCCC 234
2                CTAGGCCGGGTAGGACCCCC
3                CTAGGCCGGGTAGGACCCCC
4                CTAGGCCGGGTAGGACCCCC
5                CTAGGCCGGGTAGGACCCCC
6                CTAGGCCGGGTAGGACCCCC
7                CTAGGCCGGGTAGGACCCCC
8                CTAGGCCGGGTAGGACCCCC
9                CTAGGCCGGGTAGGACCCCC
                 ********************
```

The database exonic sequence is shown in red (from position 37 to 223) and is 187bp in size. The size of the PCR product is 234bp, which includes the exonic and the flanking intronic regions.

1 = sequence of the exon from the database (http://www.ncbi.nlm.nih/nucleotide)

2 = forward sequence 3 = reverse complement sequence of the reverse	an unaffected individual
4 = forward sequence 5 = reverse complement sequence of the reverse	affected individual from pedigree 1
6 = forward sequence 7 = reverse complement sequence of the reverse	affected individual from pedigree 2
8 = forward sequence 9 = reverse complement sequence of the reverse	affected individual from pedigree 5

Table 30 ClustalW 1.81 alignment of the exon 15 sequences of *GSY1*

```
1                                                       CCTTTCCCCTAGTACTATATGTCTGCGCG 29
2                                                       CCTTTCCCCTAGTACTATATGTCTGCGCG
3                                                       CCTTTCCCCTAGTACTATATGTCTGCGCG
4                                                       CCTTTCCCCTAGTACTATATGTCTGCGCG
5                                                       CCTTTCCCCTAGTACTATATGTCTGCGCG
6                                                       CCTTTCCCCTAGTACTATATGTCTGCGCG
7                                                       CCTTTCCCCTAGTACTATATGTCTGCGCG
8                                                       CCTTTCCCCTAGTACTATATGTCTGCGCG
9                                                       CCTTTCCCCTAGTACTATATGTCTGCGCG
                                                        *****************************

1               CCACATGGCGCTGTCCAAGGCCTTTCCAGAGCACTTCACCTACGAGCCCAACGAGGCGG 88
2               CCACATGGCGCTGTCCAAGGCCTTTCCAGAGCACTTCACCTACGAGCCCAACGAGGCGG
3               CCACATGGCGCTGTCCAAGGCCTTTCCAGAGCACTTCACCTACGAGCCCAACGAGGCGG
4               CCACATGGCGCTGTCCAAGGCCTTTCCAGAGCACTTCACCTACGAGCCCAACGAGGCGG
5               CCACATGGCGCTGTCCAAGGCCTTTCCAGAGCACTTCACCTACGAGCCCAACGAGGCGG
6               CCACATGGCGCTGTCCAAGGCCTTTCCAGAGCACTTCACCTACGAGCCCAACGAGGCGG
7               CCACATGGCGCTGTCCAAGGCCTTTCCAGAGCACTTCACCTACGAGCCCAACGAGGCGG
8               CCACATGGCGCTGTCCAAGGCCTTTCCAGAGCACTTCACCTACGAGCCCAACGAGGCGG
9               CCACATGGCGCTGTCCAAGGCCTTTCCAGAGCACTTCACCTACGAGCCCAACGAGGCGG
                ************************************************************

1               ATGCGGTGAGTGGACCCTGG 108
2               ATGCGGTGAGTGGACCCTGG
3               ATGCGGTGAGTGGACCCTGG
4               ATGCGGTGAGTGGACCCTGG
5               ATGCGGTGAGTGGACCCTGG
6               ATGCGGTGAGTGGACCCTGG
7               ATGCGGTGAGTGGACCCTGG
8               ATGCGGTGAGTGGACCCTGG
9               ATGCGGTGAGTGGACCCTGG
                ********************
```

The database exonic sequence is shown in red (from position 13 to 93) and is 81bp in size. The size of the PCR product is 108bp, which includes the exonic and the flanking intronic regions.

1 = sequence of the exon from the database (http://www.ncbi.nlm.nih/nucleotide)

2 = forward sequence
3 = reverse complement sequence of the reverse } an unaffected individual

4 = forward sequence
5 = reverse complement sequence of the reverse } affected individual from pedigree 1

6 = forward sequence
7 = reverse complement sequence of the reverse } affected individual from pedigree 2

8 = forward sequence
9 = reverse complement sequence of the reverse } affected individual from pedigree 5

Table 31 ClustalW 1.81 alignment of the exon 16B sequences of *GSY1*

```
1          TCCGCCCCACCACACTCCCCGCCTGTCCTGCCTCTCTGCTCCAGAGAGAGGATGCAGAGG 60
2          TCCGCCCCACCACACTCCCCGCCTGTCCTGCCTCTCTGCTCCAGAGAGAGGATGCAGAGG
3          TCCGCCCCACCACACTCCCCGCCTGTCCTGCCTCTCTGCTCCAGAGAGAGGATGCAGAGG
4          TCCGCCCCACCACACTCCCCGCCTGTCCTGCCTCTCTGCTCCAGAGAGAGGATGCAGAGG
5          TCCGCCCCACCACACTCCCCGCCTGTCCTGCCTCTCTGCTCCAGAGAGAGGATGCAGAGG
6          TCCGCCCCACCACACTCCCCGCCTGTCCTGCCTCTCTGCTCCAGAGAGAGGATGCAGAGG
7          TCCGCCCCACCACACTCCCCGCCTGTCCTGCCTCTCTGCTCCAGAGAGAGGATGCAGAGG
8          TCCGCCCCACCACACTCCCCGCCTGTCCTGCCTCTCTGCTCCAGAGAGAGGATGCAGAGG
9          TCCGCCCCACCACACTCCCCGCCTGTCCTGCCTCTCTGCTCCAGAGAGAGGATGCAGAGG
           ************************************************************

1          GGTGCTGCTCCTAAACCCCCGCTCCAGATCTGCACTGGGTGTGGCCCCGCAGTGCCCCCA 120
2          GGTGCTGCTCCTAAACCCCCGCTCCAGATCTGCACTGGGTGTGGCCCCGCAGTGCCCCCA
3          GGTGCTGCTCCTAAACCCCCGCTCCAGATCTGCACTGGGTGTGGCCCCGCAGTGCCCCCA
4          GGTGCTGCTCCTAAACCCCCGCTCCAGATCTGCACTGGGTGTGGCCCCGCAGTGCCCCCA
5          GGTGCTGCTCCTAAACCCCCGCTCCAGATCTGCACTGGGTGTGGCCCCGCAGTGCCCCCA
6          GGTGCTGCTCCTAAACCCCCGCTCCAGATCTGCACTGGGTGTGGCCCCGCAGTGCCCCCA
7          GGTGCTGCTCCTAAACCCCCGCTCCAGATCTGCACTGGGTGTGGCCCCGCAGTGCCCCCA
8          GGTGCTGCTCCTAAACCCCCGCTCCAGATCTGCACTGGGTGTGGCCCCGCAGTGCCCCCA
9          GGTGCTGCTCCTAAACCCCCGCTCCAGATCTGCACTGGGTGTGGCCCCGCAGTGCCCCCA
           ************************************************************

1          CCCAGTCCGCCAAACACTCCACCCCCTCCAGCTCCAGTTTCCAAGTTCCTGCACTCCAGA 180
2          CCCAGTCCGCCAAACACTCCACCCCCTCCAGCTCCAGTTTCCAAGTTCCTGCACTCCAGA
3          CCCAGTCCGCCAAACACTCCACCCCCTCCAGCTCCAGTTTCCAAGTTCCTGCACTCCAGA
4          CCCAGTCCGCCAAACACTCCACCCCCTCCAGCTCCAGTTTCCAAGTTCCTGCACTCCAGA
5          CCCAGTCCGCCAAACACTCCACCCCCTCCAGCTCCAGTTTCCAAGTTCCTGCACTCCAGA
6          CCCAGTCCGCCAAACACTCCACCCCCTCCAGCTCCAGTTTCCAAGTTCCTGCACTCCAGA
7          CCCAGTCCGCCAAACACTCCACCCCCTCCAGCTCCAGTTTCCAAGTTCCTGCACTCCAGA
8          CCCAGTCCGCCAAACACTCCACCCCCTCCAGCTCCAGTTTCCAAGTTCCTGCACTCCAGA
9          CCCAGTCCGCCAAACACTCCACCCCCTCCAGCTCCAGTTTCCAAGTTCCTGCACTCCAGA
           ************************************************************

1          ATCCACAAAGCCGTGCCTTTCTCTGGCTCCAGAATATGCATAATCAGCGCCCTGGAGTCC 240
2          ATCCACAAAGCCGTGCCTTTCTCTGGCTCCAGAATATGCATAATCAGCGCCCTGGAGTCC
3          ATCCACAAAGCCGTGCCTTTCTCTGGCTCCAGAATATGCATAATCAGCGCCCTGGAGTCC
4          ATCCACAAAGCCGTGCCTTTCTCTGGCTCCAGAATATGCATAATCAGCGCCCTGGAGTCC
5          ATCCACAAAGCCGTGCCTTTCTCTGGCTCCAGAATATGCATAATCAGCGCCCTGGAGTCC
6          ATCCACAAAGCCGTGCCTTTCTCTGGCTCCAGAATATGCATAATCAGCGCCCTGGAGTCC
7          ATCCACAAAGCCGTGCCTTTCTCTGGCTCCAGAATATGCATAATCAGCGCCCTGGAGTCC
8          ATCCACAAAGCCGTGCCTTTCTCTGGCTCCAGAATATGCATAATCAGCGCCCTGGAGTCC
9          ATCCACAAAGCCGTGCCTTTCTCTGGCTCCAGAATATGCATAATCAGCGCCCTGGAGTCC
           ************************************************************

1          CCTGGGCCTGGACCGCTTCCCAGAGGCCAGGAATCTGCCATTACTCTGCGGTGGTGCCAG 300
2          CCTGGGCCTGGACCGCTTCCCAGAGGCCAGGAATCTGCCATTACTCTGCGGTGGTGCCAG
3          CCTGGGCCTGGACCGCTTCCCAGAGGCCAGGAATCTGCCATTACTCTGCGGTGGTGCCAG
4          CCTGGGCCTGGACCGCTTCCCAGAGGCCAGGAATCTGCCATTACTCTGCGGTGGTGCCAG
5          CCTGGGCCTGGACCGCTTCCCAGAGGCCAGGAATCTGCCATTACTCTGCGGTGGTGCCAG
6          CCTGGGCCTGGACCGCTTCCCAGAGGCCAGGAATCTGCCATTACTCTGCGGTGGTGCCAG
7          CCTGGGCCTGGACCGCTTCCCAGAGGCCAGGAATCTGCCATTACTCTGCGGTGGTGCCAG
8          CCTGGGCCTGGACCGCTTCCCAGAGGCCAGGAATCTGCCATTACTCTGCGGTGGTGCCAG
9          CCTGGGCCTGGACCGCTTCCCAGAGGCCAGGAATCTGCCATTACTCTGCGGTGGTGCCAG
           ************************************************************
```

```
1          AGGTTTTAGGAAACCTGGCATGGTGCTTTCAGGTCTGGGGCTTTTAGAGCCCCCCGTGTG 360
2          AGGTTTTAGGAAACCTGGCATGGTGCTTTCAGGTCTGGGGCTTTTAGAGCCCCCCGTGTG
3          AGGTTTTAGGAAACCTGGCATGGTGCTTTCAGGTCTGGGGCTTTTAGAGCCCCCCGTGTG
4          AGGTTTTAGGAAACCTGGCATGGTGCTTTCAGGTCTGGGGCTTTTAGAGCCCCCCGTGTG
5          AGGTTTTAGGAAACCTGGCATGGTGCTTTCAGGTCTGGGGCTTTTAGAGCCCCCCGTGTG
6          AGGTTTTAGGAAACCTGGCATGGTGCTTTCAGGTCTGGGGCTTTTAGAGCCCCCCGTGTG
7          AGGTTTTAGGAAACCTGGCATGGTGCTTTCAGGTCTGGGGCTTTTAGAGCCCCCCGTGTG
8          AGGTTTTAGGAAACCTGGCATGGTGCTTTCAGGTCTGGGGCTTTTAGAGCCCCCCGTGTG
9          AGGTTTTAGGAAACCTGGCATGGTGCTTTCAGGTCTGGGGCTTTTAGAGCCCCCCGTGTG
           ************************************************************

1          GCTTACAAATTCTACAGCATACAGAGCAGGCCACGCTCAGGCCCGGCATGCGGGCCACCA 420
2          GCTTACAAATTCTACAGCATACAGAGCAGGCCACGCTCAGGCCCGGCATGCGGGCCACCA
3          GCTTACAAATTCTACAGCATACAGAGCAGGCCACGCTCAGGCCCGGCATGCGGGCCACCA
4          GCTTACAAATTCTACAGCATACAGAGCAGGCCACGCTCAGGCCCGGCATGCGGGCCACCA
5          GCTTACAAATTCTACAGCATACAGAGCAGGCCACGCTCAGGCCCGGCATGCGGGCCACCA
6          GCTTACAAATTCTACAGCATACAGAGCAGGCCACGCTCAGGCCCGGCATGCGGGCCACCA
7          GCTTACAAATTCTACAGCATACAGAGCAGGCCACGCTCAGGCCCGGCATGCGGGCCACCA
8          GCTTACAAATTCTACAGCATACAGAGCAGGCCACGCTCAGGCCCGGCATGCGGGCCACCA
9          GCTTACAAATTCTACAGCATACAGAGCAGGCCACGCTCAGGCCCGGCATGCGGGCCACCA
           ************************************************************

1          AGTTCTGGAAACCACGTGGTGT 442
2          AGTTCTGGAAACCACGTGGTGT
3          AGTTCTGGAAACCACGTGGTGT
4          AGTTCTGGAAACCACGTGGTGT
5          AGTTCTGGAAACCACGTGGTGT
6          AGTTCTGGAAACCACGTGGTGT
7          AGTTCTGGAAACCACGTGGTGT
8          AGTTCTGGAAACCACGTGGTGT
9          AGTTCTGGAAACCACGTGGTGT
           **********************
```

Exon 16 is subdivided into exons 16A, 16B, 16C and 16D. Exon 16A was screened in a previous study (Makubalo, 2000). The database exonic sequence of 16B, 16C and 16D is shown in red (from position 1 to 1110). Exonic sequence 16B (position 1 to 442bp) and is 442bp in size. The overlapping sequence of exon 16B with exon 16C is underlined.

1 = sequence of the exon from the database (http://www.ncbi.nlm.nih/nucleotide)

2 = forward sequence
3 = reverse complement sequence of the reverse } an unaffected individual

4 = forward sequence
5 = reverse complement sequence of the reverse } affected individual from pedigree 1

6 = forward sequence
7 = reverse complement sequence of the reverse } affected individual from pedigree 2

8 = forward sequence
9 = reverse complement sequence of the reverse } affected individual from pedigree 5

Table 32 ClustalW 1.81 alignment of the exon 16C sequences of *GSY1*

```
1                GGTGTCCCTGCGAATGGGGCGATCAAGTCCAGAGCCGGGGCACTTTCAGAGTTTG 492
2                GGTGTCCCTGCGAATGGGGCGATCAAGTCCAGAGCCGGGGCACTTTCAGAGTTTG
3                GGTGTCCCTGCGAATGGGGCGATCAAGTCCAGAGCCGGGGCACTTTCAGAGTTTG
4                GGTGTCCCTGCGAATGGGGCGATCAAGTCCAGAGCCGGGGCACTTTCAGAGTTTG
5                GGTGTCCCTGCGAATGGGGCGATCAAGTCCAGAGCCGGGGCACTTTCAGAGTTTG
6                GGTGTCCCTGCGAATGGGGCGATCAAGTCCAGAGCCGGGGCACTTTCAGAGTTTG
7                GGTGTCCCTGCGAATGGGGCGATCAAGTCCAGAGCCGGGGCACTTTCAGAGTTTG
8                GGTGTCCCTGCGAATGGGGCGATCAAGTCCAGAGCCGGGGCACTTTCAGAGTTTG
9                GGTGTCCCTGCGAATGGGGCGATCAAGTCCAGAGCCGGGGCACTTTCAGAGTTTG
                 ************************************************************

1                AAGGTAACTGAGAGCAGATGGTCCTCCATTTCAACTCCAGAAGTGGGGCTCTGGGA 548
2                AAGGTAACTGAGAGCAGATGGTCCTCCATTTCAACTCCAGAAGTGGGGCTCTGGGA
3                AAGGTAACTGAGAGCAGATGGTCCTCCATTTCAACTCCAGAAGTGGGGCTCTGGGA
4                AAGGTAACTGAGAGCAGATGGTCCTCCATTTCAACTCCAGAAGTGGGGCTCTGGGA
5                AAGGTAACTGAGAGCAGATGGTCCTCCATTTCAACTCCAGAAGTGGGGCTCTGGGA
6                AAGGTAACTGAGAGCAGATGGTCCTCCATTTCAACTCCAGAAGTGGGGCTCTGGGA
7                AAGGTAACTGAGAGCAGATGGTCCTCCATTTCAACTCCAGAAGTGGGGCTCTGGGA
8                AAGGTAACTGAGAGCAGATGGTCCTCCATTTCAACTCCAGAAGTGGGGCTCTGGGA
9                AAGGTAACTGAGAGCAGATGGTCCTCCATTTCAACTCCAGAAGTGGGGCTCTGGGA
                 ************************************************************

1                GGGATGTTCTAGCCCTCCCTGGCATGTCAGAGCCAGGCTCTGCCTGGAGGATCCCTCCA 607
2                GGGATGTTCTAGCCCTCCCTGGCATGTCAGAGCCAGGCTCTGCCTGGAGGATCCCTCCA
3                GGGATGTTCTAGCCCTCCCTGGCATGTCAGAGCCAGGCTCTGCCTGGAGGATCCCTCCA
4                GGGATGTTCTAGCCCTCCCTGGCATGTCAGAGCCAGGCTCTGCCTGGAGGATCCCTCCA
5                GGGATGTTCTAGCCCTCCCTGGCATGTCAGAGCCAGGCTCTGCCTGGAGGATCCCTCCA
6                GGGATGTTCTAGCCCTCCCTGGCATGTCAGAGCCAGGCTCTGCCTGGAGGATCCCTCCA
7                GGGATGTTCTAGCCCTCCCTGGCATGTCAGAGCCAGGCTCTGCCTGGAGGATCCCTCCA
8                GGGATGTTCTAGCCCTCCCTGGCATGTCAGAGCCAGGCTCTGCCTGGAGGATCCCTCCA
9                GGGATGTTCTAGCCCTCCCTGGCATGTCAGAGCCAGGCTCTGCCTGGAGGATCCCTCCA
                 ************************************************************

1                TCCGGCTCCTGTCATCCCCTACACTTTGGCCAAGCAAGAGGTGGTAGAACCACTTGGCTG 667
2                TCCGGCTCCTGTCATCCCCTACACTTTGGCCAAGCAAGAGGTGGTAGAACCACTTGGCTG
3                TCCGGCTCCTGTCATCCCCTACACTTTGGCCAAGCAAGAGGTGGTAGAACCACTTGGCTG
4                TCCGGCTCCTGTCATCCCCTACACTTTGGCCAAGCAAGAGGTGGTAGAACCACTTGGCTG
5                TCCGGCTCCTGTCATCCCCTACACTTTGGCCAAGCAAGAGGTGGTAGAACCACTTGGCTG
6                TCCGGCTCCTGTCATCCCCTACACTTTGGCCAAGCAAGAGGTGGTAGAACCACTTGGCTG
7                TCCGGCTCCTGTCATCCCCTACACTTTGGCCAAGCAAGAGGTGGTAGAACCACTTGGCTG
8                TCCGGCTCCTGTCATCCCCTACACTTTGGCCAAGCAAGAGGTGGTAGAACCACTTGGCTG
9                TCCGGCTCCTGTCATCCCCTACACTTTGGCCAAGCAAGAGGTGGTAGAACCACTTGGCTG
                 ************************************************************

1                CTCATTCCTTCTGGAGGACACACAGTCTCAGTCCAGATGCCTTCCTGTCTTTCTGGCC 725
2                CTCATTCCTTCTGGAGGACACACAGTCTCAGTCCAGATGCCTTCCTGTCTTTCTGGCC
3                CTCATTCCTTCTGGAGGACACACAGTCTCAGTCCAGATGCCTTCCTGTCTTTCTGGCC
4                CTCATTCCTTCTGGAGGACACACAGTCTCAGTCCAGATGCCTTCCTGTCTTTCTGGCC
5                CTCATTCCTTCTGGAGGACACACAGTCTCAGTCCAGATGCCTTCCTGTCTTTCTGGCC
6                CTCATTCCTTCTGGAGGACACACAGTCTCAGTCCAGATGCCTTCCTGTCTTTCTGGCC
7                CTCATTCCTTCTGGAGGACACACAGTCTCAGTCCAGATGCCTTCCTGTCTTTCTGGCC
8                CTCATTCCTTCTGGAGGACACACAGTCTCAGTCCAGATGCCTTCCTGTCTTTCTGGCC
9                CTCATTCCTTCTGGAGGACACACAGTCTCAGTCCAGATGCCTTCCTGTCTTTCTGGCC
                 ************************************************************
```

```
1          CTTTCTGGACCAGATCCTACTCTTCCTTTCTAAATCTGAGATCTCCCTCCAGG 778
2          CTTTCTGGACCAGATCCTACTCTTCCTTTCTAAATCTGAGATCTCCCTCCAGG
3          CTTTCTGGACCAGATCCTACTCTTCCTTTCTAAATCTGAGATCCTCCTCCAGG
4          CTTTCTGGACCAGATCCTACTCTTCCTTTCTAAATCTGAGATCTCCCTCCAGG
5          CTTTCTGGACCAGATCCTACTCTTCCTTTCTAAATCTGAGATCTCCCTCCAGG
6          CTTTCTGGACCAGATCCTACTCTTCCTTTCTAAATCTGAGATCTCCCTCCAGG
7          CTTTCTGGACCAGATCCTACTCTTCCTTTCTAAATCTGAGATCTCCCTCCAGG
8          CTTTCTGGACCAGATCCTACTCTTCCTTTCTAAATCTGAGATCTCCCTCCAGG
9          CTTTCTGGACCAGATCCTACTCTTCCTTTCTAAATCTGAGATCTCCTTCCAGG
           ************************************************************

1          GAATCCGCCTGCAGAGGACAGAGCTGGCTGTCTTCCCCCACCCC 822
2          GAATCCGCCTGCAGAGGACAGAGCTGGCTGTCTTCCCCCACCCC
3          GAATCCGCCTGCAGAGGACAGAGCTGGCTGTCTTCCCCCACCCC
4          GAATCCGCCTGCAGAGGACAGAGCTGGCTGTCTTCCCCCACCCC
5          GAATCCGCCTGCAGAGGACAGAGCTGGCTGTCTTCCCCCACCCC
6          GAATCCGCCTGCAGAGGACAGAGCTGGCTGTCTTCCCCCACCCC
7          GAATCCGCCTGCAGAGGACAGAGCTGGCTGTCCTCGCCCACCCC
8          GAATCCGCCTGCAGAGGACAGAGCTGGCTGTCTTCCCCCACCCC
9          GAATCCGCCTGCAGAGGACAGAGCTGGCTGTCTTCCCCCACCCC
           ********************************************
```

The database exonic sequence is shown in red. Exonic sequence 16C (position 437 to 822bp) and is 386bp in size. The overlapping sequence of exon 16B with exon 16C and exon 16C with exon 16D is underlined.

1 = sequence of the exon from the database (http://www.ncbi.nlm.nih/nucleotide)

2 = forward sequence } an unaffected individual
3 = reverse complement sequence of the reverse
4 = forward sequence } affected individual from pedigree 1
5 = reverse complement sequence of the reverse
6 = forward sequence } affected individual from pedigree 2
7 = reverse complement sequence of the reverse
8 = forward sequence } affected individual from pedigree 5
9 = reverse complement sequence of the reverse

Table 33 ClustalW 1.81 alignment of the exon 16D sequences of *GSY1*

```
1                                                    CTTCCCCCACCCCTAACCTGGCTTATTCCCAACTGCT 844
2                                                    CTTCCCCCACCCCTAACCTGGCTTATTCCCAACTGCT
3                                                    CTTCCCCCACCCCTAACCTGGCTTATTCCCAACTGCT
4                                                    CTTCCCCCACCCCTAACCTGGCTTATTCCCAACTGCT
5                                                    CTTCCCCCACCCCTAACCTGGCTTATTCCCAACTGCT
6                                                    CTTCCCCCACCCCTAACCTGGCTTATTCCCAACTGCT
7                                                    CTTCCCCCACCCCTAACCTGGCTTATTCCCAACTGCT
8                                                    CTTCCCCCACCCCTAACCTGGCTTATTCCCAACTGCT
9                                                    CTTCCCCCACCCCTAACCTGGCTTATTCCCAACTGCT
                                                     ************************************

1                  CTGCCCACTGTGAAACCACTAGGTTCTAGGTCCTGGCTTCTAGATCTGGAACCTTACCA 903
2                  CTGCCCACTGTGAAACCACTAGGTTCTAGGTCCTGGCTTCTAGATCTGGAACCTTACCA
3                  CTGCCCACTGTGAAACCACTAGGTTCTAGGTCCTGGCTTCTAGATCTGGAACCTTACCA
4                  CTGCCCACTGTGAAACCACTAGGTTCTAGGTCCTGGCTTCTAGATCTGGAACCTTACCA
5                  CTGCCCACTGTGAAACCACTAGGTTCTAGGTCCTGGCTTCTAGATCTGGAACCTTACCA
6                  CTGCCCACTGTGAAACCACTAGGTTCTAGGTCCTGGCTTCTAGATCTGGAACCTTACCA
7                  CTGCCCACTGTGAAACCACTAGGTTCTAGGTCCTGGCTTCTAGATCTGGAACCTTACCA
8                  CTGCCCACTGTGAAACCACTAGGTTCTAGGTCCTGGCTTCTAGATCTGGAACCTTACCA
9                  CTGCCCACTGTGAAACCACTAGGTTCTAGGTCCTGGCTTCTAGATCTGGAACCTTACCA
                   ************************************************************

1                  CGTTACTGCATACTGATCCCTTTCCCATGATCCAGAACTGAGGTCACTGGGTTCTAGAAC 963
2                  CGTTACTGCATACTGATCCCTTTCCCATGATCCAGAACTGAGGTCACTGGGTTCTAGAAC
3                  CGTTACTGCATACTGATCCCTTTCCCATGATCCAGAACTGAGGTCACTGGGTTCTAGAAC
4                  CGTTACTGCATACTGATCCCTTTCCCATGATCCAGAACTGAGGTCACTGGGTTCTAGAAC
5                  CGTTACTGCATACTGATCCCTTTCCCATGATCCAGAACTGAGGTCACTGGGTTCTAGAAC
6                  CGTTACTGCATACTGATCCCTTTCCCATGATCCAGAACTGAGGTCACTGGGTTCTAGAAC
7                  CGTTACTGCATACTGATCCCTTTCCCATGATCCAGAACTGAGGTCACTGGGTTCTAGAAC
8                  CGTTACTGCATACTGATCCCTTTCCCATGATCCAGAACTGAGGTCACTGGGTTCTAGAAC
9                  CGTTACTGCATACTGATCCCTTTCCCATGATCCAGAACTGAGGTCACTGGGTTCTAGAAC
                   ************************************************************

1                  CCCCACATTTACCTCGAGGCTCTTCCATCCCCAAACTGTGCCCTGCCTTCAGCTTTGGTG 1023
2                  CCCCACATTTACCTCGAGGCTCTTCCATCCCCAAACTGTGCCCTGCCTTCAGCTTTGGTG
3                  CCCCACATTTACCTCGAGGCTCTTCCATCCCCAAACTGTGCCCTGCCTTCAGCTTTGGTG
4                  CCCCACATTTACCTCGAGGCTCTTCCATCCCCAAACTGTGCCCTGCCTTCAGCTTTGGTG
5                  CCCCACATTTACCTCGAGGCTCTTCCATCCCCAAACTGTGCCCTGCCTTCAGCTTTGGTG
6                  CCCCACATTTACCTCGAGGCTCTTCCATCCCCAAACTGTGCCCTGCCTTCAGCTTTGGTG
7                  CCCCACATTTACCTCGAGGCTCTTCCATCCCCAAACTGTGCCCTGCCTTCAGCTTTGGTG
8                  CCCCACATTTACCTCGAGGCTCTTCCATCCCCAAACTGTGCCCTGCCTTCAGCTTTGGTG
9                  CCCCACATTTACCTCGAGGCTCTTCCATCCCCAAACTGTGCCCTGCCTTCAGCTTTGGTG
                   ************************************************************

1                  AAAGGGAGGGCCCCTCATGTGTGCTGTGCTGTGTCTGCACCGCTTGGTTTGCAGTTGAGA 1083
2                  AAAGGGAGGGCCCCTCATGTGTGCTGTGCTGTGTCTGCACCGCTTGGTTTGCAGTTGAGA
3                  AAAGGGAGGGCCCCTCATGTGTGCTGTGCTGTGTCTGCACCGCTTGGTTTGCAGTTGAGA
4                  AAAGGGAGGGCCCCTCATGTGTGCTGTGCTGTGTCTGCACCGCTTGGTTTGCAGTTGAGA
5                  AAAGGGAGGGCCCCTCATGTGTGCTGTGCTGTGTCTGCACCGCTTGGTTTGCAGTTGAGA
6                  AAAGGGAGGGCCCCTCATGTGTGCTGTGCTGTGTCTGCACCGCTTGGTTTGCAGTTGAGA
7                  AAAGGGAGGGCCCCTCATGTGTGCTGTGCTGTGTCTGCACCGCTTGGTTTGCAGTTGAGA
8                  AAAGGGAGGGCCCCTCATGTGTGCTGTGCTGTGTCTGCACCGCTTGGTTTGCAGTTGAGA
9                  AAAGGGAGGGCCCCTCATGTGTGCTGTGCTGTGTCTGCACCGCTTGGTTTGCAGTTGAGA
                   ************************************************************
```

```
1                GGGGAGGGCAGGAGGGGTGTGATTGGA 1110
2                GGGGAGGGCAGGAGGGGTGTGATTGGA
3                GGGGAGGGCAGGAGGGGTGTGATGGGA
4                GGGGAGGGCAGGAGGGGTGTGATTGGA
5                GGGGAGGGCAGGAGGGGTGTGATTGGA
6                GGGGAGGGCAGGAGGGGTGTGATTGGA
7                GGGGAGGGCAGGAGGGGTGTGATTGGA
8                GGGGAGGGCAGGAGGGGTGTGATTGGA
9                GGGGAGGGCAGGAGGGGTGTGATTGGA
                 ****************************
```

The database exonic sequence is shown in red. Exonic sequence 16D (position 807 to 1110bp) is 304bp in size. The overlapping sequence of exon 16C with exon 16D is underlined.

1 = sequence of the exon from the database (http://www.ncbi.nlm.nih/nucleotide)

2 = forward sequence
3 = reverse complement sequence of the reverse } an unaffected individual

4 = forward sequence
5 = reverse complement sequence of the reverse } affected individual from pedigree 1

6 = forward sequence
7 = reverse complement sequence of the reverse } affected individual from pedigree 2

8 = forward sequence
9 = reverse complement sequence of the reverse } affected individual from pedigree 5

APPENDIX IV

Contact details of collaborators:

1. L. Ashworth
   Lawrence Livermore Laboratories
   PO Box 808, L-441
   7000 East Ave
   Livermore, CA 94551
   USA
   ashworth1@llnl

2. Prof M Schalling
   Neurogenetics Unit
   Centre for Molecular Medicine
   Karolinska Institute, Sweden
   martin.schalling@cmm.ki.se

3. Prof C. Basson
   Molecular Cardiology Laboratory
   Division of Cardiology
   Department of Medicine
   Cornell University Medical College
   New York, USA
   ctbasson@mail.med.cornell.edu

4. Prof B Black
   Cardiovascular Research Institute
   Department of Biochemistry and Biophysics
   University of California
   San Francisco, USA
   bblack@itsa.ucsf.edu

5. Prof G Chandy
   Department of Physiology and Biophysics
   University of California Irvine
   Irvine
   California 92697
   USA
   gchandy@uci.edu

6. Dr P Coucke
   Department of Medical Genetics
   University of Antwerp
   Universiteitplein 1
   2610 Antwerp
   Belgium
   coucke@uia.ua.ac.be